



UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO DI  
Anglistica, Germanistica e Slavistica

DOTTORATO DI RICERCA IN  
Anglofonia

CICLO XXI

**VIDEO-INTERACTION ON *YouTube*:  
CONTEMPORARY CHANGES IN SEMIOSIS AND COMMUNICATION**

S.S.D. L-LIN/12

Coordinatore: Prof.ssa Daniela Carpi

Tutor: Prof. Cesare Gagliardi  
Prof.ssa Roberta Facchinetti  
Prof. Gunther Kress (Institute of Education, University of London)

Dottorando: Dott.ssa Elisabetta Adami



Elisabetta Adami

**VIDEO-INTERACTION ON *YOUTUBE*:  
CONTEMPORARY CHANGES IN SEMIOSIS AND  
COMMUNICATION**

*‘Why did dinosaurs evolve from water?’*  
video-interactant (reply to a comment)

*‘In nova fert animus mutates dicere formas corpora’*  
Ovid (43 BC - 18 AD), *Metamorphoses*

## ACKNOWLEDGMENTS

This work would not have achieved any coherent shape without the supervision of Gunther Kress throughout the whole research process. When I first came to him with my data, I had only a vague idea of what I was searching, let alone how to handle my ‘stuff’. Not only has he constantly encouraged me to ‘follow my hunches’, but he has been able to enlighten possible paths of interpretation out of my messy intuitions. I am especially thankful to him for all the compelling discussions which have deeply influenced my thought (to an extent which goes largely beyond the scope of the present work). His continuous raising questions (rather than providing straightforward answers) has ‘prompted’ me both to dare formulate my own answers and to learn to never stop questioning even the most banal of ‘truths’.

I am immensely thankful to Roberta Facchinetti, who has attentively read and commented the various drafts of the present work, who has supported me at every stage of the study and who has always incited me to pursue this strand of research, even if somehow ‘heretic’ within the field of linguistics. I also thank her for keeping pushing my boundaries towards things I would never dare to do otherwise.

I wish to thank Carey Jewitt, Diane Mavers and all the group at the Multimodal Research Centre at the Institute of Education. Each of them has provided insightful observations in several occasions; all of them have taught me the priceless value of sharing ideas. Thanks to the many PhD students at IoE who have willingly submitted themselves to the patient view (and review) of some of my videos. Thanks also to Silvia Bigliuzzi for her challenging feedback on a draft paper containing *in nuce* the main arguments of the present work. I owe to Mirko Grisi, my cousin and a wonderful IT ‘geek’, the design of the software for the data collection.

I deeply thank my very special family and friends for their constant support (even in my most insufferable moments) and my co-fellow student, Anna Belladelli, with whom I have shared all joys and sorrows of this three-year PhD course.

Last but definitely not least, I am grateful to all (You)Tubers, who have provided – and continue to provide – so much fascinating material for research and so many moments of pleasure, puzzlement and fun.

Elisabetta Adami, March 2009

## **ABSTRACT (English version)**

This thesis investigates the interaction by means of videos on *YouTube* Website. Video-interaction is a new form of communication which has been taking place on *YouTube* since May 2006 thanks to the introduction of the ‘video response’ option. The functionality enables (You)Tubers to reply to any given video by means of another video; hence whole communication threads are built composed of videos interacting one with another.

Given that so far no study has investigated this new type of communication, the general aim of the research is to provide a thorough description of video-interaction, in terms of both its process and products. Specifically, the analysis of the process of video-interaction focuses on (a) its distinctive features and structural characteristics, (b) its semiotic ‘affordances’ (Kress and van Leeuwen, 2001: 67), in terms of the material and social constraints and possibilities which the medium imposes over the semiosis, and (c) the diversified (and often conflicting) semiotic practices with which the affordances are actualized by the interactants. The analysis of the texts of video-interaction focuses on video-threads which start from some of the most responded videos on the Website and investigates the multimodal patterns of regularities and variations of sign-making in the chain of semiosis, that is to say, how videos establish relatedness in the thread while differentiating themselves.

The theoretical chapter reviews some of the most influential theories of communication, namely the coding-decoding and inferential models of communication (Grice, 1957, 1975; Shannon and Weaver, 1949; Sperber and Wilson, 1986), together with the notions of coherence and cohesion traditionally used in text analysis (Beaugrande and Dressler, 1981; Fairclough, 1992; Halliday and Hasan, 1976; van Dijk, 1985). Furthermore, it confronts these models and notions with the practices of video-interaction; finally, it discusses the inadequacies of these theories for the description of video-interaction, crucially because, in video-interaction, the interlocutors’ mutual understanding of their intended meaning is not essential for communication to succeed. On these grounds the framework adopted for the analysis is introduced, i.e., social semiotics multimodal analysis (Hodge and Kress, 1988; Kress and van Leeuwen, 1996, 2006; Kress and van Leeuwen, 2001). Within this framework and on the basis of the social-semiotic category of ‘interest’ (Kress and van Leeuwen, 1996, 2006: 13), the heuristic notion of an ‘interest-driven prompt-response relation’ is devised as an analytic tool used for the description of both the process and the texts of video-interaction.

The methodological chapter discusses the issues of representativeness, significance, reproducibility and verifiability implied in collecting a corpus of online data. It illustrates the criterion of popularity which has driven the selection of the data in order to overcome the aforesaid hardly solvable issues. A review of

the current practices of transcription highlights their inaptness for the purposes of the present research and motivates the *ad hoc* transcription devised for the threads. Then the chapter illustrates the analytical methodology, which, in a cyclic process, has involved all stages of the research, from the selection of the data and their transcription, to the pilot study and up to the consequent refinement of the theoretical framework and of its analytic tools. The analysis follows a funnel process; indeed, from the regularities and variations detected at more general levels, it zooms in to more fine-grained levels of analysis. The analysis combines quantitative and qualitative methods with a textual interpretation focused on signifiers (on the semiotic resources present in the texts, rather than on their signifieds). Finally, the chapter discusses the ethical stance which has grounded the choice of conducting a covert observation on the Website, with no prior consent asked to the participants. This choice is motivated by the manifest publicity of the Website (and by the criterion of popularity driving the data selection) and by the intention of avoiding any patronizing attitude towards the authors of the videos, considered here as film-makers. This standpoint adds to the debate currently ongoing on online research ethics.

The first analysis chapter focuses on the process of video-interaction, by examining its structural characteristics, the affordances of the video response option and the practices with which these affordances are exploited by interactants. The analysis of the structural characteristics of video-interaction leads to the mapping of its place within our contemporary semiotic landscape. The study of the affordances leads to a comparison between what is possible to do with the video response option and how these possibilities are used creatively by the interactants, according to their interests. In turn, these creative uses often lead to implementations to the medium and to changes in its affordances. Through a 14-month period of observation of the Most Responded Videos top chart on the Website, the analysis singles out a taxonomy of the types of most responded videos, identified in (a) video requests, (b) prompting videos, (c) anomalous most responded videos, and (d) flooded related-response videos. The chapter concludes by describing the distinctive practices for each typology.

Moreover, the so-identified taxonomy enables the research to analyse the texts of video-interaction, through the selection of an exemplary instance for each typology of video-threads. The analysis of the texts of video interaction – detailed in two separate chapters – is based on almost two thousand videos. By focusing on the largest instances of video-interaction currently existing, the analysis can identify the multimodal patterns of sign-making in interaction, in terms of regularities and variations (rather than of ‘rules’ and ‘exceptions’, given that video-interaction is a new form of communication in which conventions are not established yet, but are rather being continuously developed, transformed and negotiated, sometimes in conflicting ways).

Rather than following traditionally conceived cooperative or relevance principles and notions of coherence and cohesion, video-interaction works on a 'loose' form of individualized participation. While the structural characteristics determine the interactional possibilities, the interactants' practices exploit the affordances of video-interaction in unexpected ways and often lead to changes in the structure itself (in the same way as the Saussurean *parole*, when made socially dominant, modifies the *langue*). Patterns of relatedness in the exchanges are driven by the sign-makers' diversified interests. More specifically, sign-making relations in the exchanges are established through a system of differentiation-within-attuning, so that interactants bring their distinctive contribution while keeping the same (either formal or semantic) theme, set by the initial video. In this sense, video-interaction is analogous to collective forms of artistic improvisation (e.g., the genre of variation in music or the contemporary free-style). Video responses take up – and actualize – one or more prompts within the range of possibilities set by the initial video. This prompt-response relation generally disregards the interlocutor's intended meaning, it often does not follow relevance principles, it presents (traditionally considered) marked textual organizations, and does not build coherent exchanges. Not only is (intertextual) implicitness widely practiced, but also formal attuning is sometimes more regarded than (semantic) coherence for the establishment of relatedness in the exchange.

Rather than on the interactants' mutual understanding, video-interaction hinges on a playful and challenging engagement with the medium and with the other texts in the exchange, by exploiting maximally the representational possibilities of both. Patterns of relatedness are established through an interest-driven selection, transformation, assemblage and recontextualization of (often formal, at times implicit or backgrounded) elements of the initial video or of other texts. This way, sign-making is often done by means of a copy-and-paste technique, which follows the interactants' diversified interests, and disregards coherence or the sign-maker's intended meaning. Responses frequently and intentionally produce their texts through misunderstanding and exploit maximally the potential semantic ambiguity of the interlocutor's text. Incoherent exchanges are successful and generally acknowledged and accepted by interactants. In this sense, video-interaction epitomizes the changes in representation and communication which are taking place in our contemporary semiotic landscape, whereby traditional systems of coherence are disregarded and (traditionally considered) incoherent or non-cohesive exchanges are acceptable as the result of representations produced through the selection, copy and paste, recontextualization and forwarding of other texts.

These contemporary changes in representation and communication need analogous changes in the theories of communication, while their description requires adequate analytical models. Therefore the thesis concludes by hypothesizing the usefulness of the theoretical and analytical framework devised for the research to the description of contemporary forms of communication.

At an empirical level, the original contribution of this thesis consists in providing a description of a new form of communication. At a theoretical level, it highlights the inadequacies of traditional theories of communication to the description of contemporary forms of communication and attempts at devising new analytical tools which may be more apt to the task.



## ABSTRACT (Italian version)

La tesi ha per oggetto la ‘video-interazione’, ossia una nuova forma di comunicazione che ha luogo sul sito *YouTube* grazie all’introduzione – nel maggio 2006 – dell’opzione di ‘video risposta’, tramite la quale un video può fungere da risposta ad un altro video. Questa funzionalità consente agli interagenti di costruire scambi comunicativi attraverso dei video. Il suo impiego genera intere catene comunicative, costituite appunto da video che rispondono l’un l’altro.

In considerazione dell’assenza di studi su questo nuovo tipo di comunicazione, la ricerca mira a fornire una descrizione accurata della video-interazione, sia in termini di processi che di prodotti. Nello specifico, l’analisi del processo si focalizza su (a) i tratti distintivi e le caratteristiche strutturali della video-interazione in quanto forma di comunicazione, (b) le ‘*affordances*’ semiotiche (Kress and van Leeuwen, 2001: 67), in termini di ciò che il mezzo consente o impedisce (e promuove o stigmatizza) sia a livello materiale (tecnologico) che di convenzioni sociali, e (c) le pratiche semiotiche diversificate (e spesso conflittuali) secondo cui le *affordances* vengono attualizzate dagli interagenti. D’altra parte, l’analisi dei testi della video-interazione s’incentra su *video-threads* (filoni d’interazione video), che prendono avvio dai video che hanno ricevuto il maggior numero di video risposte e esamina i *patterns* multimodali – in termini di regolarità e di variazione – dei processi di significazione nella catena della semiosi, cioè le modalità con cui le video risposte si relazionano al video iniziale e tra loro nel filone.

Il capitolo teorico rivisita alcune delle più influenti teorie di comunicazione, quali i modelli comunicativi di codifica-decodifica (Shannon and Weaver, 1949) e quelli inferenziali (Grice, 1957, 1975; Sperber and Wilson, 1986), insieme alle nozioni di coerenza e coesione tradizionalmente utilizzate nell’analisi testuale (Beaugrande and Dressler, 1981; Fairclough, 1992; Halliday and Hasan, 1976; van Dijk, 1985). Mediante un confronto con le pratiche semiotiche in uso nella video-interazione, il capitolo evidenzia le inadeguatezze di tali teorie per la descrizione della video-interazione, essenzialmente in ragione del fatto che, in quest’ultima, la reciproca comprensione del significato intenzionale degli interagenti non è essenziale perché scambi comunicativi di successo abbiano luogo. In considerazione di ciò, viene presentato il quadro di riferimento adottato per l’analisi, ovvero l’analisi multimodale socio-semiotica (Hodge and Kress, 1988; Kress and van Leeuwen, 1996, 2006; Kress and van Leeuwen, 2001). All’interno di tale quadro e sulla base della nozione socio-semiotica di ‘interesse’ (Kress and van Leeuwen, 1996, 2006: 13), lo studio introduce l’euristico di ‘relazione di *prompt-response*’ dettata dall’interesse del *sign-maker* (segnificatore). Tale euristico, derivato dall’osservazione stessa delle pratiche di significazione nei filoni d’interazione video, viene adottato come strumento analitico e descrittivo sia del processo che dei testi della video-interazione.

Il capitolo metodologico discute delle problematiche della raccolta di dati online, in termini di rappresentatività del *corpus* e di significatività, riproducibilità e verificabilità dei risultati, e illustra il criterio di popolarità che – per ovviare a tali problematiche – ha guidato la selezione dei dati. Una riesamina dei metodi di trascrizione esistenti ne evidenzia l’inutilizzabilità per gli scopi del presente lavoro e motiva la trascrizione *ad hoc* formulata per i testi del *corpus*. Successivamente il capitolo illustra la metodologia d’analisi, che ha coinvolto in maniera ciclica ogni stadio della ricerca, dalla selezione dei dati, alla loro trascrizione, allo studio pilota e alla conseguente messa a punto del quadro teorico di riferimento e degli strumenti analitici. L’analisi segue un processo ad ‘imbuto’, che, dall’identificazione di regolarità e ‘eccezioni’ ai livelli più generali, si focalizza su livelli d’analisi sempre più dettagliati. L’analisi integra metodi di tipo quantitativo e qualitativo con l’interpretazione testuale incentrata sui significanti (sulle risorse semiotiche presenti nei testi piuttosto che sui significati). Il capitolo discute, infine, la posizione etica alla base della scelta di una metodologia di osservazione nascosta delle pratiche in atto sul sito, senza la previa richiesta di consenso ai partecipanti. Tale scelta è motivata dall’esplicito status pubblico del sito (e dal criterio di popolarità che ha guidato la selezione dei dati) e dalla volontà di evitare atteggiamenti paternalistici nei confronti degli autori dei video, considerati qui come veri e propri *film-makers*. Tale presa di posizione apporta nuovi contributi all’acceso dibattito attualmente in corso sull’etica della ricerca online.

Il primo capitolo d’analisi tratta il processo della video-interazione, attraverso l’esamina delle caratteristiche strutturali del mezzo, delle *affordances* dell’opzione della video risposta e delle pratiche con cui queste vengono sfruttate dagli interagenti. Le caratteristiche strutturali della video-interazione consentono la mappatura del posto che questa occupa all’interno del panorama semiotico contemporaneo. La disamina delle *affordances* della funzionalità della video risposta consente un raffronto tra ciò che è possibile fare attraverso di essa e come tali possibilità vengano sfruttate in maniera creativa dagli interagenti. Tali usi creativi portano spesso a implementazioni al mezzo stesso e alle sue *affordances*. Attraverso un periodo di osservazione di 14 mesi della classifica dei video più risposti sul sito, l’analisi ricava una tassonomia dei video che danno il via ai filoni più numerosi, identificati in (a) video richieste, (b) video ‘stimolatori’ (*prompting*), (c) video ‘anomali’ e (d) video ‘inondati’ da risposte pertinenti. Per ciascuna tipologia vengono analizzate le pratiche semiotiche distintive in atto.

Tale tassonomia consente inoltre l’analisi dei testi della video-interazione, attraverso la selezione di un esemplare per ciascuna tipologia di filoni video. L’analisi dei testi della video-interazione – dettagliata in due capitoli separati – si basa su un *corpus* di quasi due mila video. Focalizzandosi sugli scambi comunicativi più estesi attualmente esistenti (i filoni più numerosi), l’analisi identifica i *patterns* multimodali di significazione in interazione, in termini di regolarità e di variazioni attualizzate (piuttosto che di ‘norme’ e ‘eccezioni’,

trattandosi di una forma di comunicazione del tutto nuova in cui le convenzioni non sono ancora normalizzate, bensì in continuo sviluppo, trasformazione e negoziazione, talvolta conflittuale).

Piuttosto che sui tradizionali principi cooperativi e di rilevanza, o sulle nozioni di coerenza e di coesione, la video-interazione si basa su una 'blanda' forma di partecipazione individuale. Mentre le caratteristiche strutturali del mezzo determinano le possibilità comunicative, le pratiche semiotiche sviluppate dagli interagenti sfruttano le *affordances* della video-interazione in modi inediti, che spesso conducono a cambiamenti alla struttura stessa (analogamente al saussuriano atto di *parole* che, divenuto sociale e dominante, incide sulla *langue*). Le modalità con cui i video si relazionano l'un l'altro nelle interazioni sono il risultato degli interessi diversificati dei significatori. Più precisamente, le relazioni di significazione negli scambi seguono un sistema di differenziazione e assonanza, o 'accordatura' (*attuning*), cosicché gli interagenti apportano il proprio contributo distintivo mantenendo al contempo un tema (formale o semantico) comune, dettato dal video iniziale, in modo analogo a forme collettive d'improvvisazione artistica (es. dal genere della 'variazione' in musica, all'odierno *free-style*). Le video risposte colgono – e realizzano – uno o più stimoli (*prompts*) all'interno della gamma delle possibilità stabilite dal video iniziale. In molti casi, tale relazione di *prompt-response* non considera il significato intenzionale dell'interlocutore, spesso non segue principi di rilevanza, presenta strutture informative tradizionalmente considerate marcate e non costruisce scambi coerenti. Non solo l'implicitezza (spesso intertestuale) è largamente praticata, ma anche l'assonanza formale è talvolta privilegiata rispetto alla coerenza (semantica).

Piuttosto che sulla comprensione reciproca degli interagenti, la video-interazione si basa su un giocoso e provocatorio confronto col mezzo e con gli altri testi dello scambio, sfruttando al massimo le possibilità semiotiche di entrambi. I testi si relazionano l'un l'altro negli scambi comunicativi attraverso la selezione interessata, la trasformazione, l'assemblaggio e la ricontestualizzazione di elementi (spesso formali, a volte impliciti o non salienti) del testo a cui rispondono o di altri testi. In tal modo la significazione è spesso prodotta tramite tecniche di 'copia e incolla' che seguono gli interessi diversificati degli interagenti, ignorano le tradizionali regole di coerenza e il significato intenzionale dell'interlocutore, facendo invece ampio uso del fraintendimento volontario e sfruttando al massimo la potenziale ambiguità semantica del testo dell'interlocutore. Scambi del tutto incoerenti hanno successo comunicativo e sono generalmente accettati e approvati dagli interagenti. In tal senso, la video-interazione è emblematica dei mutamenti in corso nei sistemi di comunicazione contemporanei, in cui i tradizionali sistemi di coerenza vengono a mancare e scambi tradizionalmente considerati non coesi o incoerenti sono perfettamente accettabili, in quanto risultanti da rappresentazioni prodotte tramite selezione, copia-incolla, ricontestualizzazione e inoltro di altri testi.

Tali mutamenti del panorama semiotico contemporaneo necessitano di teorie comunicative e modelli di analisi adeguati. Pertanto la tesi conclude ipotizzando l'utilità del quadro teorico e analitico sviluppato dallo studio per la descrizione delle forme di comunicazione contemporanea.

A livello empirico, il contributo originale della tesi consiste nel fornire la descrizione di una nuova forma di comunicazione. A livello teorico, il lavoro mette in luce le inadeguatezze delle teorie di comunicazione tradizionali per la descrizione di forme di comunicazione contemporanea e tenta di sviluppare nuovi strumenti analitici più adatti a tale compito.

# CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>4</b>
<b>ABSTRACT (English version).....</b>	<b>5</b>
<b>ABSTRACT (Italian version) .....</b>	<b>9</b>
<b>CONTENTS .....</b>	<b>13</b>
<b>CHAPTER 1. INTRODUCTION.....</b>	<b>19</b>
<b>CHAPTER 2. THEORETICAL FRAMEWORK .....</b>	<b>29</b>
1 MODELS AND THEORIES OF COMMUNICATION.....	30
1.1 The coding/decoding model.....	30
1.2 The Gricean inferential model .....	30
1.3 Relevance Theory .....	31
1.4 The models confronted with video-interaction.....	33
2 THE TEXTUAL NOTION OF COHERENCE (AND COHESION).....	35
2.1 Coherence confronted with video-interaction.....	38
3 SOCIAL SEMIOTICS .....	40
3.1 Contemporary social (semiotic) changes .....	40
3.2 Relevant concepts and categories .....	43
3.3 The notion of affordances .....	47
3.4 The notion of interest.....	48
4 CATEGORIES USED IN THE ANALYSIS.....	49
4.1 Video-interaction as process and video-interaction as texts.....	49
4.1.1 The prompt-response relation in the process .....	50
4.1.2 The prompt-response relation in the texts.....	52
4.1.3 The analyses of texts and processes: responded prompts .....	53
4.1.4 Caveat: the prompt-response relation in a social semiotic perspective .....	54
4.2 The participants in video-interaction .....	56
4.3 The protagonists of video-interaction .....	57
4.4 The texts of video-interaction .....	58
4.4.1 Texts within texts.....	58
4.5 Further terminology: Video-interaction, (video-)thread and (You)Tuber .....	59
5 CONCLUSIONS.....	61

<b>CHAPTER 3. METHODOLOGY .....</b>	<b>63</b>
1 DATA RETRIEVED ONLINE: THORNY ISSUES.....	63
1.1 Representativeness and significance .....	64
1.2 Stability, reproducibility and verifiability .....	64
1.2.1 Non-storability of YouTube materials.....	65
2 DATA SELECTION.....	66
2.1 The criterion of popularity in video-interaction as process.....	67
2.1.1 The monitoring period of the process.....	67
2.2 The criterion of popularity in video-interaction as texts .....	68
2.2.1 The time-span of the collection of the texts .....	70
2.2.2 The corpus of texts .....	70
3 THE TRANSCRIPTION.....	72
3.1 Transcription practices for multimodal materials.....	73
3.1.1 Dynamic images .....	74
3.2 The <i>ad hoc</i> transcription devised for the data.....	76
3.2.1 An interest-driven selective transduction .....	76
3.2.2 A cyclic process driven by recurrence, saliency and relevance .....	76
3.2.3 The transcribed (recurrent, salient and relevant) features .....	77
3.2.4 A ‘monomodal’ transcription of multimodal texts .....	78
4 THE METHOD OF ANALYSIS.....	78
4.1 The funnel process.....	79
4.2 The pilot study .....	79
4.3 Quantitative and qualitative interpretation focused on signifiers.....	80
5 ETHICS .....	81
5.1 Ethical stance in conducting covert observation .....	82
5.2 Ethical stance in the presentation of the data .....	84
5.3 Copyright issues .....	85
6 CONCLUSIONS .....	86
<b>CHAPTER 4. ANALYSIS 1/3: VIDEO-INTERACTION AS PROCESS .....</b>	<b>89</b>
1 STRUCTURE: DISTINCTIVE FEATURES .....	89
1.1 (Embodied and disembodied) multimodality .....	90
1.2 Homogeneity and bidirectionality .....	95
1.3 Publicity.....	95
1.4 Asynchronicity .....	97
1.5 Disembodiment.....	97
1.6 Online .....	98
1.7 Distance .....	98

1.8	Multiple mediation.....	99
1.9	Corporate interface distribution.....	100
1.10	The place of video-interaction in the semiotic landscape.....	102
2	THE ‘VIDEO RESPONSE’ OPTION: AFFORDANCES AND PRACTICES .....	103
2.1	The Video Response Option.....	104
2.2	Related videos vs. video responses.....	104
2.3	Technical affordances .....	107
2.4	The affordances in use .....	109
2.4.1	The approval/denial and the power of the initiator.....	110
2.4.2	The denial and the process of interaction .....	110
2.4.3	Responses creation and thread composition .....	111
2.4.4	Textual organization as clue for responses creation .....	112
2.4.5	The ‘Play all responses’ option and the thread as an entity .....	114
2.4.6	The ‘Play all responses’ option and the sublevels in the thread .....	115
2.4.7	The sequential display and the values of ‘firstness’ and ‘newness’ .....	116
3	THE ‘MOST RESPONDED’ TOP CHART: AFFORDANCES AND PRACTICES.....	116
3.1	Type of Most responded videos.....	117
3.1.1	Video requests.....	118
3.1.2	Prompting videos .....	121
3.1.3	Anomalous most responded videos .....	124
3.1.4	Related flooding responses .....	126
4	CONCLUSIONS.....	129
<b>CHAPTER 5. ANALYSIS 2/3: ‘WHERE DO YOUTUBE?’ THREAD .....</b>		<b>133</b>
1	THE THREAD COMPOSITION.....	134
2	THE INITIAL VIDEO.....	135
3	THE VIDEO RESPONSES.....	139
3.1	Topic relatedness: The answer to the question .....	141
3.1.1	The geographical location.....	142
3.1.2	The non-geographical location .....	144
3.1.3	The prompt-response continuum of the represented location.....	145
3.1.4	The signifiers of the location .....	150
3.2	Formal relatedness: The (marked) mode as attuning device .....	155
3.3	Textual organization: The relatedness-continuum.....	160
3.3.1	Narrative structure .....	161
3.3.2	Unmarked organization: Location as New and focus.....	162
3.3.3	Marked organization: Location as Given and circumstantial element .....	163
3.3.4	Cohesive ties in the paratext .....	165
3.3.5	No (explicit) clues of relatedness.....	168

3.3.6 The relatedness-continuum: to sum up .....	172
4 SUB-RESPONSES.....	175
4.1 Sub-responses answering the topic-question.....	175
4.2 Self-responses: topic specification, development and diversion.....	176
4.3 No clues of relatedness .....	177
5 THE VIDEO-SUMMARY .....	178
6 THE RESPONSES TO THE SUMMARY .....	184
7 CONCLUSIONS .....	185
<b>CHAPTER 6. ANALYSIS 3/3: ‘BEST VIDEO EVER!’ THREAD .....</b>	<b>189</b>
1 THE <i>CHRIS CROCKER</i> PHENOMENON .....	189
1.1 The semiotic activity around the initiator.....	190
2 THE INITIAL VIDEO .....	193
2.1 The ideational and interpersonal meaning of the video.....	195
2.2 The further meaning of the paratext .....	196
2.3 A possible range of interpersonal prompts .....	197
3 THE VIDEO RESPONSES .....	198
3.1 Corresponding responses.....	200
3.1.1 The blink as a corresponded prompt.....	200
3.1.2 The ‘insignificant action’ as a corresponded prompt .....	203
3.1.3 The sexually-related corresponded prompt .....	206
3.1.4 The corresponded smile.....	206
3.1.5 The ‘bestness’ as the corresponded prompt.....	207
3.1.6 Variations in the corresponded prompts .....	208
3.2 Commentary responses.....	211
3.2.1 The initial video commented .....	212
3.2.2 CC’s character as the commented prompt.....	215
3.2.3 The comment as a chance for further representations .....	219
3.3 Remakes: From remix to parody .....	222
3.3.1 Recontextualization of the initial video.....	222
3.3.2 Recontextualization of other CC’s videos .....	228
3.4 Original spoofs .....	232
3.4.1 Spoofs of the initial video.....	233
3.4.2 Spoofs of CC’s character.....	233
3.5 Inferential responses .....	239
3.5.1 Inferential prompts from the initial video.....	240
3.5.2 Inferential prompts from the character .....	241
3.6 Responses with secondary reference to CC.....	248



3.7	Random-related responses .....	251
3.8	Paratext-related responses.....	252
3.9	Unrelated responses .....	254
4	SUB-RESPONSES.....	257
5	CONCLUSIONS.....	259
<b>CHAPTER 7. CONCLUSIONS .....</b>		<b>263</b>
1	ACHIEVEMENTS OF THE RESEARCH .....	263
1.1	Theoretical and empirical achievements.....	263
1.2	Methodological achievements .....	266
2	LIMITATIONS OF THE RESEARCH .....	267
3	FUTURE DEVELOPMENT .....	268
<b>REFERENCES .....</b>		<b>269</b>



# CHAPTER 1

## INTRODUCTION

*‘e mi chiesi se questo che mi chiude  
ogni senso di te, schermo d’immagini’*

E. Montale, *Mottetto* (1938)

Pictures are worth a thousand words, as is commonly said. Even more so today, in the era of Web 2.0, when home-made pictures can virtually ‘move’, join dynamically language and sound, and reach every (connected) corner of the world.

This is probably the reason why *YouTube*<sup>1</sup>, the leading Website for uploading, sharing and viewing video clips, is daily accessed by millions of people and is currently one of the most visited<sup>2</sup> Websites of the whole World-Wide-Web. This is also the motive that drives mass media to pay such huge attention to it; indeed, while conducting the present research, I have personally subscribed the ‘News Alert’ service offered by *Google* search engine for the keyword *YouTube* and, for the last two years, the daily email posted to my account has never recorded less than 10 entries (which is the maximum number provided by this service), and this accounts only for news sources published online in English.

To the same extent that *YouTube* is dealt with in the news, news circulates massively on *YouTube*, as evidenced by the war fought by means of videos promoting or ‘spoofing’ the campaigners for the 2008 US presidential elections. The political activity – among all the others – which takes place on *YouTube* is so influential that, at the moment of writing, an interdisciplinary academic conference is being organized, which focuses on ‘YouTube and the 2008 Election Cycle in the United States’ (University of Massachusetts Amherst, April 16 - 17, 2009, Amherst, MA)<sup>3</sup>.

As the above cited conference exemplifies, in recent times, academic research in various disciplines has started to investigate the *YouTube* phenomenon. Notwithstanding the very young age of the Website (created in February 2005) and the sometimes long time required for academic papers to be publicly available, a substantial number of works are currently being published in computer science and information technology (Baluja et al., 2008; Benevenuto et al., 2008a; Benevenuto et

---

<sup>1</sup> [www.youtube.com](http://www.youtube.com)

<sup>2</sup> In February 2009, Alexa ([www.alexa.com](http://www.alexa.com)) ranks *YouTube* as the third most visited Website, after two search engines (*Google* and *Yahoo!*). (Retrieved 26 February 2009).

<sup>3</sup> For details, cf. <http://www.umass.edu/polsci/youtube> (Retrieved 4 February 2009).

al., 2008b; 2007: 114; Capra et al., 2008; Cattuto et al., 2007; Cha et al., 2007; Cheng et al., 2007; Duarte et al., 2007; Gill et al., 2007; Halvey and Keane, 2007; Jain, 2007; O'Donnell et al., 2008; Paolillo, 2008; Ulges et al., 2008; Weiss, 2007; Xia et al., 2007; Yahia et al., 2007; Zink et al., 2008), in education and learning (Berlanga et al., 2007; Cann, 2007; Duffy, 2007; Eastment, 2007; Freitas et al., 2008; Gromik, 2007; Jenkins, 2007; Sébastien et al., 2007; Trier, 2007a, b; Webb, 2007), in sociology (Clemons et al., 2007; Godwin-Jones, 2007; Halvey and Keane, 2007), in computer-mediated communication (Barnes and Hair, 2007; Lange, 2007a, b, 2008), and in ethnographic and cultural studies (Bardzell, 2007; Burgess, forthcoming; Burgess and Green, 2008; Carroll, 2008; Melican and Faulkner, 2007; Molyneaux et al., 2008; Regan and Revels, 2007; Shida and Gater, 2007; Willett, forthcoming), among others (Bruns et al., 2007; Carlson et al., 2008; Cohen and Küpçü, 2007; Gueorguieva, 2008; Keen, 2008; Kessler, 2007; Lee, 2008; Lewis, 2007; Madden, 2007; O'Brien, 2007; Pace, 2008; Turkheimer, 2007). Eventually, digital ethnography is investigating *YouTube* by using its very same medium<sup>4</sup>. All these works – the list is increasing every day – try to shed light on the online video-sharing phenomenon by adopting different theoretical perspectives and methodological/analytical approaches and by focusing on different aspects of the activity which takes place on this very popular Website.

My research aims at doing quite the same, by focussing on a distinctive type of communication that takes place on *YouTube*. Indeed, a recent enhancement of the *YouTube* interface, the 'video response' option (introduced in May 2006), enables '(You)Tubers' to post videos in reply to any video uploaded on the Website. Thanks to this feature, (You)Tubers can now *interact* by means of videos, so that whole communication threads are created through videos addressing one to another.

The interaction by means of video clips – or, as I call it, video-interaction – is a brand new communication practice. By means of a video, one can not only communicate, but also reply to someone else's video, through virtually all semiotic modes available: through gestures and facial expressions, spoken and written language, sounds and music, drawings and animations, filming and photos; only body contact is excluded among interactants. Indeed, in semiotic terms, videos allow for both 'embodied' and 'disembodied' modes (Norris, 2004. Cf. also Chapter 4, section 1.1) to be deployed through space and time to make meaning in a disembodied text (i.e., a semiotic artefact, separated from its producer, which can be replayed at will by the viewer).

When videos – which in themselves are highly multimodal communicative units – are used in interaction, they give a unique opportunity for the observation of planned and asynchronous multimodal communication and, ultimately, of human communication as a whole. Indeed, the already complex intertwining of semiotic

---

<sup>4</sup> Cf., the very promising ethnographic investigation currently being carried out by M. Wesch and his digital ethnographic working group at Kansas State University, retrievable online at: <http://mediatedcultures.net/youtube.htm> and at <http://mediatedcultures.net/ksudigg/>

resources deployed in videos develops in even more complex ways when these resources make meaning in interaction.

Communication always involves interaction. As Kress and van Leeuwen put it,

by communicating we interact, we do something to or for or with people – entertain them with stories, persuade them to do or think something, debate issues with them, tell them what to do, and so on. (Kress and van Leeuwen, 2001: 114)

In Speech Act Theory terms (Austin, 1962; Searle, 1969), every communication instance is simultaneously a locutionary, illocutionary and perlocutionary act, i.e., it says something (its content), it does something (states, invites, asks, etc.) towards someone, in order to get something (persuade, convince, provoke, etc.) done by them. When communication occurs in an interactional exchange (as in the case of a video replying to another) the form and content of the locutionary act are shaped not only by the intended illocutionary act (i.e., what the text is intended to perform) and the intended perlocutionary effect (i.e., what the text is intended to achieve), but also by the form and content of the previous turn in the exchange (i.e., the video-as-text to which we reply).

Interaction is a mutually influencing active relationship. So, for example, when we reply to somebody's email, the electronic text that we write and send is influenced both by what we are replying to and by the effect(s) we want our reply to have on our addressee. This 'prompt-response' relation shapes both content and form of our text, whose very existence lies in its being meant to be a communicative unit in and for an exchange<sup>5</sup>. When interaction is made by means of videos – in which a wider range of modes can be deployed than in, e.g., emails – the observable influences and effects of their being a communicative unit in interaction are multiplied, essentially because the semiotic resources available to (be prompted and responded by) interactants are in turn multiplied.

This wide range of representational possibilities in interaction constitutes the first reason which has driven me to investigate this new form of communication. Specifically, I wanted to observe how videos select their representational resources among all the possible ones and how this selection is influenced by the other texts in the interaction. This has given rise to the first (chronologically conceived) research question of this work, namely:

1. Which are the multimodal patterns of regularities and variations in the texts of video-interaction; that is to say, how does each video relate to the others in the interaction?

At any given time, the ways in which we communicate are shaped not only by the

---

<sup>5</sup> This is the reason why no communicative unit in interaction can be accounted for its meaning without (at least) putting it into relationship with its other co-interacting units.

specific interaction within which our semiotic act takes place, but also by the material and social ‘affordances’ (Kress and van Leeuwen, 2001: 67) of the media which we use to design, produce and distribute our texts<sup>6</sup>. Indeed every medium enables (or fosters) and prevents (or hinders) certain communicative practices, both materially (e.g., emails allow for asynchronous communication while they prevent face-to-face real time interaction) and socially (e.g., the socially established functions of emails are different from those of – say – text messaging). Therefore, while starting to approach the phenomenon, I have increasingly been interested in investigating the affordances and the practices which are being developed in this new form of communication. In video-interaction, the affordances of the video response option on the interface – what it enables and prevents interactants to do – shape the practices used by interactants. In turn, these practices (how the affordances are used by interactants), can – and indeed do – (re)shape the affordances of the medium and, by so doing, they can lead to changes in the whole structure (as much as the Saussurean *parole* draws on and realizes the *langue*, which, in turn, is incessantly reshaped by culturally and socially determined yet individual acts of *parole*).

This second motive of my investigation (affordances and practices) has led me to broaden the research focus from the products of video-interaction (i.e., the texts) to the process of communication instantiated in them. This has given rise to a second research question, namely:

2. What are the affordances of video-interaction and how are they actualised in the interactants’ practices?

At a theoretical level, the first strand of my research (i.e., the observation of multimodal patterns in the texts of video-interaction) has led me to broaden my reference framework from linguistics to semiotics; indeed, as anticipated, the analysis of language can account only partially for the meaning which is made in videos through the deployment of a wide range of semiotic resources. The second strand (i.e., the observation of the practices of video-interaction), has led me to further broaden my reference framework from a multimodal analysis of the texts to the theories and models of communication which could be more apt to describe the practices which I was observing.

In sum, the aim of my research is to investigate (a) the affordances of video-interaction and (b) how they are exploited and reshaped by its participants, in both (1) the practices and (2) the texts. In other words, I am interested in how participants ‘cope’ with the rules of the structure (i.e., constraints and possibilities) and affect them, while they ‘tune in’ with others in the interaction and, at the same time, bring their own specific contribution to it.

---

<sup>6</sup> For design, production and distribution as the three main processes of semiosis, cf. Kress and van Leeuwen (2001).

As said, this general research aim involves a two-fold task, which the description of the present work rearranges as follows:

1. the investigation of the process of video-interaction;
2. the investigation of its products, i.e., the texts in the actual exchanges.

In sum, I am interested in identifying the characteristics of video-interaction in this early stage of its development, with a particular focus on (a) the affordances of the structure and the interactional practices with which it is used, and on (b) the multimodal patterns of regularity and variation that emerge in the video-threads which constitute the texts of the interaction.

To do so, I adopt a social semiotic perspective (Halliday, 1978; Hodge and Kress, 1988) to investigate the video response option and its use, as actualized in the largest interactive instances currently available (those started by the videos which appear on the ‘Most Responded of All Time’ top chart on *YouTube*). This analysis enables the outlining of the distinctive characteristics of the process of video-interaction. On the basis of this, the texts of video-interaction are investigated by means of a multimodal analysis (Kress and van Leeuwen, 1996, 2006) on video-threads which start from some of these most responded videos of ‘All Time’. The analysis of the texts enables the outlining of the distinctive patterns with which videos establish relatedness within the interaction.

Ultimately, the whole work tries to adopt (and adapt) Kress and van Leeuwen’s (2001) ‘multimodal theory of communication’ – which, in their words, means investigating

two things: (1) the *semiotic resources* of communication, the modes and the media used, and (2) the communicative practices in which these resources are used.  
(2001: 111)

The contribution of the present study is – I believe – both empirical and theoretical. At the empirical level, the research provides the description of a new type of interaction, which has not been thoroughly investigated yet. At a theoretical level, it confronts traditional communication theories and models with this contemporary form of communication. Indeed, as will be seen, video-interaction defies traditional notions of ‘cooperation’ (Grice, 1967, 1975), ‘relevance’ (Sperber and Wilson, 1986), ‘coherence’ and ‘cohesion’ (Halliday and Hasan, 1976), in that, rather than on the participants’ mutual understanding, successful communication in video-interaction works on a highly differentiated-and-attuned system of prompt-response relations driven by the interactants’ diversified interests. Rather than the coherence of the exchange, it is the interested selection, recontextualization, transformation and assemblage of other people’s texts which determines successful communication. On this basis, the observation of video-interaction provides theoretical insights on how contemporary forms of interaction ‘deviate’ from traditional models of communication, and, thus, on how new heuristics and models are needed to give a

thorough account of contemporary forms of communication and semiosis.

In order to introduce the theoretical framework of the research, Chapter 2 reviews some of the main theories and models of communication, from the coding-decoding ones, such as Shannon and Weaver's model (Shannon and Weaver, 1949), to the inferential ones, such as Grice's conversational maxims (Grice, 1957, 1967) and Sperber and Wilson's Relevance Principle (1986), together with the traditional textual analysis notions of 'coherence' and 'cohesion' (Halliday and Hasan, 1976). As anticipated, these theories, models and notions prove themselves inadequate to describe and explain thoroughly video-interaction. On these grounds, the rationale of the theoretical framework used to approach the subject is discussed, by introducing some basic concepts and categories of Social Semiotics (Halliday, 1978; Hodge and Kress, 1988), whose distinctive take on communication makes it more apt to handle both the process and the products of video-interaction. The discussion focuses particularly on the notions of 'affordance' and 'interest', then it introduces the heuristic notion of an interest-driven 'prompt-response relation', which has been devised for the analysis of both the process of video-interaction and of its texts. The chapter concludes by defining the main categories and the terminology used in the analysis.

Chapter 3 illustrates the methodology adopted for the research, in terms of data selection, transcription and analysis, as well as the ethical stance of the research. Firstly, it discusses some thorny issues implied in selecting online data, namely those of representativeness and significance, of stability, reproducibility and verifiability, together with the problematic non-storability of *YouTube* data. Secondly, it illustrates the corpus selected for the analysis, both by discussing the criteria adopted for its selection and by describing the type of data. The criterion of popularity has driven the selection of the data used for both the analysis of the process of video-interaction and of its texts. The analysis of the process has produced a taxonomy of the most responded videos – which start the largest instances of video-interaction currently existing – which, in turn, has led to the selection of the threads used for the analysis of the texts. Thirdly, the chapter discusses the issues involved in the transcription of dynamic materials (i.e., videos) in interaction; it reviews some existing transcription practices for multimodal data and for dynamic images, and it introduces the rationale for the *ad hoc* transcription that has been devised for the data. Fourthly, the chapter describes the various steps of the analytical method. Following a funnel process, the analysis has played a major role in every stage of the research, from the data selection to their transcription, from the pilot study to the validation of its results onto the whole corpus; moreover, in a cyclic process, it has led to the (re)formulation and refinement of the theoretical framework and analytic tools. By focussing on signifiers, the analysis combines qualitative and quantitative methods, so as to give scientific evidence to the interpretation of the data. Finally, the chapter steps into the very sensitive fields of ethics, privacy and copyright policies, inevitably implied in this type of research. Indeed, since videos are the primary data of the study, personal information concerning the identities of the interactants are hardly concealable in the



description, if the research wants to account for how meaning is made with a multiplicity of modes, including, for example, gestures and facial expressions. Hence, the chapter concludes with the ethical position endorsed here for the treatment of sensitive information that may be disclosed in the data description.

Chapter 4 opens to the first strand of data analysis, by focusing on the process of video-interaction. Firstly, it analyses the structural characteristics of video-interaction, by singling out its distinctive features, so as to map out the place of video-interaction within the contemporary semiotic landscape. After considering the structure, the analysis examines the introduction and the affordances of the video response option, in terms of what it enables interactants to do, both materially and socially. Then, it focuses on the introduction, the affordances and the (varied) use of the ‘Most Responded Videos’ top chart. It examines the types of videos which have been charted over a 14-month monitoring period, thus leading to the further selection and analysis of the texts. The analysis of the process of video-interaction gives evidence to the fact that the affordances of the medium influence the way it is used in interaction, while, simultaneously, the practices – with which affordances are exploited according to the interactants’ (diversified and often conflicting) interests – lead to modifications in the affordances themselves.

Chapter 5 starts the analysis of the texts. It is devoted to the analysis of a video-thread started by a topic-specific request, i.e., the video titled ‘@----Where Do YouTube?----@’. It illustrates the thread composition, it describes the initial video and the possible range of prompts it sets. It then analyses the video responses, according to the different ways in which they establish relatedness with the initial video and to the different ways in which they organize their representations, thus actualizing various prompt-response relations in the thread. The same is done with the sub-responses (videos posted as responses to the initial video’s responses), before describing and analysing the video-summary (which has been made by the thread initiator through a selection and recontextualization of shots of the responses) and the responses to the summary. The analysis of this thread evidences to a trend of attuning in form, together with a high differentiation within this attuning. It further testifies to notable exceptions to traditional notions of cooperation, relevance and coherence, such as instances of maximum exploitation of semantic ambiguity, (traditionally considered) marked textual organizations, and totally unrelated exchanges considered as successful by the participants.

The ‘exceptional’ deviations observed in Chapter 5 become the ‘rule’ in the thread analysed in Chapter 6, i.e., the ‘Best Video EVER!’ video-thread, which is started by a ‘prompting’ video (rather than by a topic-specific request). Here again, after the illustration of the thread composition, of the character of the thread initiator, and of the initial video, the analysis is organized according to the type of relatedness that the video responses (and sub-responses) establish with the initial video. The analysis of this video-thread testifies to the fact that video-interaction works on a diversified and individualized participation (rather than on Gricean ‘cooperation’), and that

representation is achieved through an interested selection, recontextualization and transformation of resources, rather than being driven by coherence. What noticed in the previous thread becomes manifest in this one, i.e., that the initial video sets the ground for the other representations; it determines the possible range of prompts to which videos can respond; in turn, responses can take up and make salient even the remotest element of (or implied in) the initial video, without ever impeding successful communication to take place. This is mainly due to the fact that here successful communication disregards traditional communicative principles such as ‘the interlocutors’ mutual understanding’ and is rather determined by an individualized participation in the semiotic space driven by – and fulfilling – the interactants’ diversified interests.

Finally, Chapter 7 summarizes the results of the analysis. It draws some conclusions concerning the main theoretical, empirical and methodological achievements of the research; it further mentions some limitations of the study and a possible agenda for future research in the field. It attempts also some generalizations of the results in terms of possible patterns and practices which are distinctive of contemporary forms of communication and interaction.

Pictures are worth a thousand words. I have chosen to start my thesis with this common place because it is appropriate not only for the object of my study, but also for its description. Throughout the work, I attempt at giving written descriptions of the visual and auditory representations that are in videos; I try to force my (limited) linguistic resources and this (strict) academic genre as much as I can, to give *literally* the *sense* of drawings, gestures, facial expressions, sounds, hand-writing, typographical layouts and filming effects. I endeavour myself to make descriptions as simple, neat and precise as possible, and yet I must admit that they are very often rather complex, vague and unfit to describe what I have been watching in these videos. Indeed, even the shortest and most trivial video needs several paragraphs of written description, and still some meaning is inevitably lost in the ‘transduction’ (Kress, 2003) from one mode to another – or, rather, from the high and dense multimodality of videos to the very constrained modality of academic writing. As Bateson puts it,

[T]he messages that we exchange in gestures are really not the same as any translation of those gestures into words (Bateson, 1953: 129)

This is maximally true when a gesture is provided with colour effects, subtitles, soundtracks and ‘emoticons’. To somehow compensate the meaning lost in transduction, a substantial part of the description relies on snapshots of videos; yet snapshots cannot account for the meaning that is made through the deployment through time of the semiotic resources in videos.

In sum, if images are worth a thousand words, dynamic images (i.e., videos) are worth a thousand images and, possibly, one hundred thousand words (like those of

this thesis). Again, it is probably the reason why *YouTube* is now among the most visited Website of the whole World-Wide-Web, and this is definitely the reason why I have been totally fascinated by these sometimes trivial videos interacting with each other. Ultimately, this is also the reason why this research would be more apt to be delivered in a multimedia format, to give a better idea – in a more effective, possibly shorter and certainly more entertaining way – of the characteristics of the interaction that takes place on *YouTube*.

Fortunately, I can rely confidently on two factors that can fill what my inadequate description lacks. Firstly, I can rely on some shared knowledge of the semiotic resources that I am describing, since, each day, our lives are massively surrounded by videos (and, even when not making and watching videos, we all indeed interact in a highly multimodal way), so that, by experience, *we all know what it's all about*. Secondly, to use a spatial metaphor very common to (You)Tubers, *that bulk of amazing stuff is all out there!* And one just needs to browse and click on the primary source of my data, to *replay* and make one's own meaning out of it. I warmly recommend it.



## CHAPTER 2

### THEORETICAL FRAMEWORK

*‘C'est par le malentendu universel que tout le monde  
s'accorde. Car si, par malheur, on se comprenait, on  
ne pourrait jamais s'accorder’*

P. C. Baudelaire, *Mon cœur mis à nu* (1864)

Linguists and philosophers of language have widely debated on the principles which explain human communication and interaction. As described by Sperber and Wilson (1986: 1-37), throughout the last century, models of coding-decoding (e.g., Shannon and Weaver, 1949) have been paired (and contrasted) with inferential approaches to communication (Grice, 1967; 1975; Sperber and Wilson, 1986).

Although both these models can account for many types of communication, while approaching the data of the present work, they have proved to be hardly apt for the analysis of video-interaction, which seems to work on rather different conventions and interactional processes, thus shaping in new ways traditional notions of cooperation and relevance. Also coherence, a widely acknowledged semantic principle used in text and discourse analysis (Beaugrande and Dressler, 1981; Fairclough, 1992; Halliday and Hasan, 1976; van Dijk, 1985), is differently actualized – and often disregarded – in video-interactional exchanges.

The present chapter discusses briefly these theories and notions and confronts them with the peculiarities of video-interaction (Section 1 and Section 2). Then the chapter introduces the theoretical framework adopted for the present analysis, i.e., social semiotics (Halliday, 1978; Hodge and Kress, 1988; van Leeuwen, 2005), whose perspective allows for a more thorough account of video-interaction (Section 3). Within this general framework, the heuristic notion of an interest-driven prompt-response relation (Adami, 2009c) is introduced, which has been devised so as to describe and explain the specific sign-making patterns of video-interaction. Finally, the chapter concludes detailing the analytical tools adopted here, together with a short definition of the key-concepts used for the analysis of the data (Section 4).

# 1 MODELS AND THEORIES OF COMMUNICATION

## 1.1 The coding/decoding model

A widely known coding-decoding model is Shannon and Weaver's (1949), according to which a message is coded by the sender and transmitted through a channel to the receiver, who decodes it. For communication to be successful, the sender and the receiver must share the same code and the channel must not be disturbed (e.g., by noises), so that the message is transmitted correctly from the source to the target through this coding-decoding process. The epistemology underlying this communication model is rather linear; one meaning/concept is associated to (encoded into) one form/signal and has to be recovered by the receiver through decoding. At one end, signification is a matter of a correct application of coding rules and, at the other end, meaning-making is a matter of a correct application of decoding rules.

## 1.2 The Gricean inferential model

Theories always reflect the social conditions of their times<sup>7</sup>, so, when, in Western societies, traditional power relationships were being shaken, alternative models of communication were devised which questioned the linearity of the coding-decoding model and the power attributed to the source in favour of a greater role assigned to the receiver. Philosophy of language began to undermine the linear metaphor of transmission of one meaning from a source to a target. Indeed, while Austin's (1962) Speech Acts Theory began to conceive language in use for what it did rather than (just) said, Grice (1957; 1967; 1975) investigated meaning which is implied rather than said and the inferences that the hearer/reader is to do in order to understand the meaning of an utterance.

In general terms, Grice can be grouped with Austin, Searle, and the later Wittgenstein as 'theorists of communication-intention' (Miller, 1998: 223; Strawson, 1971: 172). The belief of this group is that intention/speaker-meaning is the central concept in communication, and that sentence-meaning can be explained (at least in part) in terms of it. (Davies, 2007: 2317)

As is well known, Grice (1967; 1975) listed four maxims – Quantity, Quality, Relation and Manner – which drive the hearer/reader's inferences in understanding the meaning of an utterance out of the many possible ones, on the basis of an assumed 'cooperative principle' to which communicators generally conform, unless stated/proved otherwise. This assumed cooperative principle goes as follows:

---

<sup>7</sup> 'not only semiotic resources and their uses but also theories and methods arise from the interests and needs of society at a given time, whether researchers are aware of this or not' (van Leeuwen, 2005: 69).

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (Grice, 1975: 26)

Unlike the coding/decoding model of communication, the inferential one conceives any utterance as having more than one possible meaning; the ‘correct’ one is inferred by virtue of assuming that a cooperative principle is at work in communication, i.e., that the speaker says neither more nor less than what she<sup>8</sup> believes to be (a) necessary, (b) true and (c) relevant, (d) in as a clear and ordered manner as possible<sup>9</sup>. When ‘what is said’ does not conform to this principle, ‘implicatures’ are generated to derive an implied meaning which satisfies the maxims.

### 1.3 Relevance Theory

Sperber and Wilson’s Relevance Theory (1986) brings Grice’s model a step further. They reduce Grice’s maxims to one, i.e., relevance, and assign the meaning of an utterance entirely to the interpretative activity of the hearer/reader, who interprets every text on the basis of hypotheses on the relevance that the communicative event has on a given situation. They undermine the notion of shared/mutual knowledge needed for the inferential process to be successful, in favour of a ‘weaker’ notion of ‘mutual manifestation’ to the interlocutors of relevant assumptions which drive the interpretation. These are selected according to the principle of aiming at achieving the maximum cognitive effect (in terms of information interpreted out of an utterance) with the minimum effort (in terms of inferential work). According to this principle, rather than given, the (relevant) context is selected and searched for at any given time in the interaction.

In spite of these differences, Sperber and Wilson still rely on Grice in conceiving successful communication in terms of the hearer’s understanding of the speaker’s intentions, so that they use the inferential process based on relevance to determine successful and unsuccessful communication, understood as a process of correct/incorrect interpretation of the speaker’s intended meaning.

Undoubtedly, Grice’s inferential model of communication and Sperber and Wilson’s Relevance Theory are ‘looser’ approaches to communication than the coding/decoding one. They take into consideration the so-called ‘extra linguistic’ factors, such as the context, and the possibility that a given utterance – understood in a wider sense by Sperber and Wilson, so that it includes also non-linguistic semiotic acts – may be interpreted in different ways in different contexts and by different interlocutors.

---

<sup>8</sup> The present work selects ‘generic *she*’ for generic reference; cf. the findings in Adami (2009b).

<sup>9</sup> These correspond to the Gricean maxims of (a) quantity, (b) quality, (c) relation and (d) manner.

However, as anticipated, both Grice's and Sperber and Wilson's models conceive successful communication in terms of correct retrieval/interpretation of the meaning intended by the speaker. Indeed, 'Grice places [the emphasis] on the role of speaker-intention in the process of meaning-recognition' (Davies, 2007: 2319). Consequently,

as a hearer, I should recognise why you said something and any change in my beliefs should come (at least in part) from what is said. Communication is thus characterised as an active process where a speaker (or communicator) attempts to convey their belief to the hearer. (Davies, 2007: 2319)

Analogously, although in Sperber and Wilson's view, 'communication is governed by a less-than-perfect heuristic' (1986: 45), they nevertheless agree with Grice in stating that '[c]ommunication is successful [...] when hearers [...] infer the speaker's "meaning"' (1986: 23), otherwise communication fails:

the process of inferential comprehension is non-demonstrative: even under the best of circumstances [...] communication may fail. The addressee can neither decode nor deduce the communicator's communicative intention. The best he can do is construct an assumption on the basis of the evidence provided by the communicator's ostensive behaviour. (1986: 65)

Many subsequent works have criticized, questioned, implemented, proved or refuted either one or the other model, or both<sup>10</sup>, yet they have never questioned their shared main assumption, i.e. that communication is a matter of understanding the utterer's intended meaning.

In this sense, Blakemore's interpretation of Sperber and Wilson's theory is maybe the most relativistic one:

the representation that the audience derives [...] should not be seen as a copy or literal representation of the communicator's thought, but as an interpretation of it –

---

<sup>10</sup> The literature on both Grice's and Relevance Theory is huge (Ariel, 2002; Bezuidenhout and Cooper Cutting, 2002; Bird, 1979; Breheny, 2006; Fredsted, 1998; Gibbs, 1987; Giora, 1997, 1999; Levinson, 1989; Mey and Talbot, 1988; Sadock, 1986; Zhang, 1998). Gricean theories have been commented and revised by several scholars (Atlas, 1989, 2005; Atlas and Levinson, 1981; Bach, 1999a, b, 2001a, 2002a, b; Benz and van Rooij, 2007; Berg, 1991; Bird, 1979; Brumark, 2006; Burt, 2002; Capone, 2006; Davis, 1998; Davis, 2007; Delgrande et al., 2005; Gazdar, 1979; Gazdar and Good, 1982; Harnish, 1976; Haugh, 2002; Horn, 1985, 2004; Leech, 1983; Levinson, 1983; Myers-Scotton, 1993; Neale, 1992; Récanati, 2002, 2004; Sadock, 1978; Saul, 2002; Ziv, 1988). Some have proposed substantial modifications, such as the notion of 'implicature' (Bach, 1994a, b, 2001a, b), or 'presupposition' (Strawson, 1971), which more or less open to interpretative ambiguity (Chierchia and McConnell-Ginet, 2000; Stalnaker, 1999). In general, many conflicting readings and interpretations of Grice have been given (Lindblom, 2001); for a recent survey, cf. Horn (1985) and for a recent re-reading of Grice (critical of Sperber and Wilson), cf. Davies (2007). Sperber and Wilson's theory has been variously introduced (Blakemore, 1995; Wilson, 1999), revised, supported (Blakemore, 1987, 2002; Blass, 1990; Carston, 1999a, b, 2002a, b; Stainton, 1998; Wilson and Sperber, 2004), and criticized (Clark, 1987; Cooren and Sanders, 2002; Kasher, 1991; Levinson, 2000). Some works try to give reason of both theories (Båve, 2008; Chien, 2008; Colston, 2000).



that is, as a representation that resembles the communicator's thought in virtue of sharing its logical and contextual implications. (Blakemore, 2008: 42)

However, in her understanding of 'resemblance' as utterer's and audience's representations giving 'rise to the same logical and contextual implications' (Blakemore, 2008: 44), she nevertheless associates successful communication with the interpreter's (sufficiently faithful) 'recovery' of the communicator's meaning:

successful communication is achieved when a communicator produces a public representation of one of his/her thoughts about a state of affairs and the audience recovers a representation that is a sufficiently faithful interpretation of that thought. (Blakemore, 2008: 45)

In sum, no one has ever questioned the rather commonsensical and intuitive idea that successful interaction is a matter of interactants understanding each other (i.e., the intended meanings behind their semiotic acts)<sup>11</sup>.

Also in the wide debate (cf. Capone, 2006 for an extensive review) between what is to be assigned to semantics and what to pragmatics, pragmatists always refer to the speaker's intended meaning:

There is, I claimed, no such thing as 'what the sentence says' in the literalist sense, that is, no such thing as a complete proposition autonomously determined by the rules of the language with respect to the context but independent of the speaker's meaning [...] in order to reach a complete proposition through a sentence, we must appeal to the speaker's meaning. (Récanati, 2004: 59)

So, even when truth values between propositions and world, and univocal relations between meaning and form have been denied, still speakers' intentions are unquestionably critical to successful communication.

#### **1.4 The models confronted with video-interaction**

When applied to video-interaction, these theories of communication reveal themselves to be unsuitable to describe the patterns with which videos reply one to another. As will become evident in the analysis (Chapters 4, 5 and 6), video-interaction is a very 'loose' form of communication, even looser than the one explained by Relevance Theory<sup>12</sup>.

---

<sup>11</sup> Roland Barthes has challenged radically the idea in his *The Death of the Author* (1977a), which is one of the main sources of Social Semiotics, the theoretical perspective adopted here (cf. Section 3).

<sup>12</sup> In fact, Sperber and Wilson distinguish between 'strong' and 'weak forms of communication'; in the latter, vagueness is an intrinsic feature of meaning and 'the communicator can merely expect to stir the thoughts of the audience in a certain direction' (1986: 60). Yet they ascribe weak communication typically to non-verbal forms. The supposed polysemy of images compared to

Indeed, in video-interactional exchanges, it is seldom what the author means in her video that counts, but rather, what the would-be respondent *can mean out of* the representation of the former. Rather than what *you* intended to mean in your video, it is what *I*, according to my interests, can make of your video which drives me to respond to it with another video. In sum, in video-interaction, texts (i.e., videos) are conceived as semiotic resources to be transformed, reused, (mis)interpreted and misquoted in and for other texts.

This generally acknowledged, aware and pursued misinterpretation does not make video-interaction a non-cooperative or unsuccessful type of communication. Indeed, here successful communication is not a matter of a correct interpretation of the interlocutor's intended meaning; it is rather a matter of taking up one or more prompts of her text and respond to it in a creative way. Cooperation is often at place (together with conflict) in video-interaction but it is a rather different one from Grice's understood principle. Here interaction deals with using and transforming other people's representations for new ones, so that 'cooperation' (if we are to keep this label) is an acknowledged form of 'interested contribution' and follow-up; it is a rather 'loose' form of participation, in terms of an interested productive response to a prompt.

If 'Grice is concerned with the distinction between saying and meaning – how hearers recognize the utterer's intention when speakers use implicit language' (Davies, 2007: 2313), video-interaction can not be explained in terms of right or wrong hypotheses/inferences of the speaker's intentions out of the many possible ones, but rather in terms of an interested response out of a range of possible prompted ones (or, even better, an interested response which realizes a prompt out of a given ground of possibilities). This by no means entails that inferential processes are not at work in interpreting the sign-maker's intentions, nor that codes and conventions do not exist and are not negotiated among participants. However, both the inferential process of understanding the sign-maker's intended meaning and the existence of codes and conventions at work cannot alone account for the sign-making which is made in video-interaction. Were it the case, too many instances would need to be considered as exceptions, as a fruit of a disturbed channel (which is not likely to occur much often, due to the little technological requirement of *YouTube* video streaming) or of a non-shared code on the one hand, or of a disregard of Gricean maxims and relevance principles on the other.

True, it could be maintained simply that video-interaction is a genre which conventionally breaks Gricean maxims so as to achieve rhetorical effects such as irony, for example<sup>13</sup>. As is well known, irony (i.e., to mean the opposite of what is

---

verbal language has been claimed also by Barthes (1977b: 32-51); for a convincing argument against this view, cf. Kress and van Leeuwen (1996, 2006: 17-18).

<sup>13</sup> In this regard, cf. Colton's (2000) 'ironic implicature hypothesis'; for a different perspective on irony see Clift (1999). Giora and others have focused on how poetic devices such as repetitions, analogies and irony impair discourse comprehension (Giora, 1990, 1993, 1995; Giora et al., 1996).

said) is not the only case in which the cooperative principle is disregarded:

Grice nowhere says, nor would want to say, that all conversations are governed by the co-operative maxims. There are too many garden-variety counter-examples; social talk between enemies, diplomatic encounters, police interrogation of a reluctant suspect, most political speeches, and many presidential news conferences. These are just some of the cases in which the maxims of co-operation are not in effect, and are known not to be in effect by the participants. (Harnish, 1976: 340, note 29).

However, as will be seen, unlike the ‘garden-variety counter-examples’ cited by Harnish, video-interaction cannot be said to be non-cooperative and participants are not (always) comparable to enemies, reluctant suspects or political opponents, in that the goals of their ‘manipulative’ strategies, even if sometimes responding to challenges, are not to deceive or defeat/win over each other.

Besides, even if some instances of video-interaction can be indeed assimilated to a sort of playful contest<sup>14</sup> in which manipulation – yet deprived of any negative connotation – is one of the acknowledged conventions, it still remains the rather puzzling situation in which too many exchanges seem a dialogue between interactants who are deaf not only to what is ‘said’ (i.e., represented in each other’s videos) but also to each other’s intended meaning, as if produced by some of Beckett’s protagonists. In other words, even if approved by the interactants<sup>15</sup> (and thus, to some extent at least, ‘successful’), many video-interaction exchanges, seem totally ‘incoherent’ (i.e., they are composed of texts which give no clues to be related). This latter observation calls into question the notion of ‘coherence’, which is reviewed briefly in the following section in order to determine to what extent it can be adopted in the analysis of video-interaction.

## 2 THE TEXTUAL NOTION OF COHERENCE (AND COHESION)

‘As Grice has noted, utterances need to be related in order to make up a coherent and cooperative conversation’ (Capone, 2006: 650). More or less coeval to Gricean theories, the notion of ‘coherence’ arose in functional linguistics, started by Halliday and Hasan’s (1976) work on cohesion and then variously taken up by studies in text linguistics (Beaugrande and Dressler, 1981) and by (Critical) Discourse Analysis (Fairclough, 1992; van Dijk, 1979; 1980a; 1980b; 1985).

Indeed, in 1976, Halliday and Hasan define ‘cohesion’ as a semantic concept which ‘refers to relations of meaning that exist within the text, and that defines it as a text’

---

Cf. also Costall and Leudar’s (2007) mention of ‘irony, sarcasm, deception, lies, persuasion and other manipulative uses of language’.

<sup>14</sup> The contest is a well established genre of video-interaction (cf. also Chapter 4, Section 3.1.1).

<sup>15</sup> Indeed the video uploader has the power to accept or refuse any video response to her video (cf. the following Section 2, on ‘coherence’; cf. also Chapter 4, Sections 2.3 and 2.4.1).

(1976: 4). Cohesion is a 'semantic relation' (1976: 5) of the text expressed by means of ties in the text's lexicon and grammar, which drive the reader through interpreting the text as a 'coherent' whole. In their analysis of cohesive ties, they include also interactional exchanges (uttered by two or more interlocutors), so that this extends the notion of cohesion – and its related one of coherence – also to the meaning which is made in interaction; so does van Dijk: 'the various coherence constraints also hold for dialogical discourse' (1985: 121).

If cohesion is 'the set of possibilities that exist in the language for making text hang together' (1976: 18),

the texture involves more than the presence of semantic relations of the kind we refer to as cohesive [...]. It involves also some degree of coherence in the actual meanings expressed: not only, or even mainly, in the CONTENT, but in the TOTAL selection from the semantic resources of the language, including the various interpersonal (socialexpressive-conative) components - the moods, modalities, intensities, and other forms of the speaker's intrusion into the speech situation [...]. A text is a passage of discourse which is coherent in these two regards: it is coherent with respect to the context of situation, and therefore consistent in register; and it is coherent with respect to itself, and therefore cohesive. (1976: 23)

Also Halliday and Hasan's coherence depends to some extent on the interlocutor's interpretative work:

Texture is a matter of degree. It is almost impossible to construct a verbal sequence which has no texture at all – but this, in turn, is largely because we insist on interpreting any passage as text if there is the remotest possibility of doing so. We assume, in other words, that this is what language is for. (1976: 23)

Robert de Beaugrande and Wolfgang Dressler ascribe coherence as the second of the two text-centred standards, among their 7 principles of texture, and define it as 'the mutual access and relevance within a configuration of concepts and relations' underlying the text surface (1981).

According to van Dijk, 'the proposition sequence underlying an acceptable discourse must satisfy various conditions of what is called coherence' (1985: 108). In various works, van Dijk (1979; 1980a; 1980b) distinguishes between 'local' and 'global' coherence. The former is referred to the logical connections in the sequence of propositions in a text<sup>16</sup> and is signalled in the surface structure of the text by means of devices of cohesion (in Halliday and Hasan's terms), while the latter ('global coherence') refers to a topic or theme of a discourse, i.e., the semantic information that provides the overall unity of a text:

[A] discourse also has a global semantic structure or macrostructure (Jones, 1977; van Dijk, 1972; 1977; 1980a). Thus a macrostructure is a theoretical reconstruction

---

<sup>16</sup> i.e., 'what is discovered first should be mentioned first', '[p]ossible, probable, or necessary conditions (e.g., causes) should in general be mentioned before their consequences' (1985: 109).

of intuitive notions such as topic or theme of a discourse. It explains what is most relevant, important, or prominent in the semantic information of the discourse as a whole. At the same time, the macrostructure of a discourse defines its global coherence.’ (1985: 115)

Van Dijk hedges his definition of coherence by mentioning the possibility of

intentional deviations having specific functions (esthetic, as in literary discourse; didactic, as in examples; or rhetorical and stylistic deviation or variation for special purposes). (1985: 122).

Furthermore he admits that coherence is variously realized according to different genres: ‘discourses that do not allow summarizing have no macrostructure or only a very fragmentary one (e.g., some modern poems)’ (1985: 117). However, ‘[w]ithout a semantic macrostructure, even a fragmentary one, there is no overall coherence and hence no point to the discourse’ (1985: 131).

Apart from different degrees of coherence in different genres, coherence assignment varies also according to different participants:

since the world knowledge, beliefs, opinions, attitudes, interests, and goals of speech participants may vary, they may also assign different global meanings (macrostructures) to the same discourse as they may have different evaluations about what is relevant or important information for the discourse (and the communicative context) as a whole. (1985: 117)

Nevertheless, ‘[d]espite these individual and subjective variations, there is often enough overlap to guarantee successful communication and interaction’ (1985: 117), i.e., again, for interactants to understand each other.

Relying on van Dijk’s (1979) definition of coherence and criticizing Sperber and Wilson’s theory, Giora (1997)<sup>17</sup> argues that ‘[s]peakers and hearers may have other goals than just enriching their cognitive environment or that of their addressees’ (1997: 31), and that Sperber and Wilson’s relevance,

cannot be the only principle that governs human communication. Speakers and hearers are not constrained only by the search for relevance. In addition, coherence considerations constrain communication and play a major role in discourse structuring and understanding. (1997: 31)

Here, Giora argument is that coherence is relevant to communication since it is relevant, once again, to discourse ‘understanding’.

In sum, also for the theorists of coherence, successful communication and interaction is a matter of mutual understanding – i.e., the participants’ assignment of a

---

<sup>17</sup> For other works on coherence, which also include discussions of Grice’s inferential model, cf. among others Bezuidenhout (2006), Craig and Tracy (1983) and Tomlin (1987).

substantially mutual global meaning.

## 2.1 Coherence confronted with video-interaction

Fairclough (1992) subjectivizes even more the notion of coherence, defining it as an essential aspect of interpretation:

coherence is not a property of texts, but a property which interpreters impose upon texts, with different interpreters possibly generating different coherent readings of the text. (1992: 133)

Even if more relative and open to possibilities, i.e., multiple readings, also this notion of coherence seems to be too tight to describe video-interaction (or, at least, some of its instances). Indeed, according to Fairclough ‘interpreters reduce the potential ambivalence of texts’ (1992: 81); this ambivalence reduction is achieved through processing both the ‘context’ conceived ‘in a narrow sense [...] as that which preceeds (or follows) in a text’ (1992: 81) and the ‘context of situation’, so that

these interpretations lead to predictions about the meanings of texts which again reduce ambivalence by excluding certain otherwise possible meanings. (1992: 81)

However, in video-interaction, rather than reducing what Fairclough calls ‘ambivalence’, interactants strive to exploit it. Indeed, not only many videos – and often the most intriguing ones – defy any notion of coherence in themselves (i.e., within a single text, which often seems an ‘arbitrary’ mashup of random signs), but many video responses seem to defy it on purpose, by exploiting even the remotest ambiguity of the text, conceived as a chance for making a new, mostly unexpected, often distorted and unrelated (i.e., incoherent) meaning out of the videos they respond to, so as to differentiate themselves, catch the viewers’ attention and puzzle them.

Just to make an example, how could one define the topic/theme (van Dijk’s ‘global coherence’) of a 4” video featuring a (You)Tuber’s face smiling and blinking at the camera?<sup>18</sup> And, more significantly, even if a somewhat generic global topic were assigned to this text, such as ‘blinking at the camera’ (which, however would not say anything about the ‘intended’ meaning of this video), how could one determine its relatedness to one of its video responses (i.e., the coherence of the whole interactional exchange) which, two minute-long, features a chest-bared (You)Tuber dancing and pouting in front of the camera? Yet, it would be wrong to conclude that this is an instance of unsuccessful communication, essentially for two reasons:

- firstly, because this type of exchanges is too frequent in the corpus (so that

---

<sup>18</sup> The example is taken from the initial video ‘Best Video EVER!’ and one of its responses in the thread analysed in Chapter 6.

- one should conclude that video-interaction is generally unsuccessful);
- secondly, because the author of the initial video has the power to reject any response attempt to her video, so that the very presence of these unrelated videos linked as responses to the initial one is the proof that the interaction is, at least to some extent, ‘successful’.

Therefore, if we are to maintain the traditional models of communication and the notion of coherence as they are, either we are forced to (1) conclude that participants in these interactional exchanges draw on knowledge which is not shared by the viewers to understand each other, or else, paradoxically, we are forced to (2) exclude video-interaction from the realm of human communication, which is to say that what happens on *YouTube* is not at all interaction, that no one communicates anything to anyone. Of course, none of these options is tenable, considering, on the one hand, the millions who enjoy watching these video-interaction instances (and hence make some sense out of them, even if they do not share the participants’ background knowledge) and, on the other, the thousands who participate in this apparently incoherent, yet successful, type of interaction, by uploading video responses to other videos (so that some kind of communication must necessarily take place there).

It must be noted that, as mentioned for the coding-decoding and the inferential models of communication, we are not here discarding the notion of coherence as a whole just because video-interaction does not conform to it in a traditional way. Indeed, if, as said above, one of the goal of responding to a video with a topic-unrelated one may reside in puzzling the viewers (so as to catch their attention), this ‘puzzling’ effect is achieved precisely because the viewers’ expectations of a coherent exchange are not fulfilled. In this sense, paradoxically, posting an incoherent response is pragmatically coherent, i.e., it is coherent with the goal of the interaction, which, in this case, is to produce puzzlement in the viewer.

Indeed, Van Dijk (1980b) describes pragmatic (functional) coherence as distinct from semantic coherence:

One of the prominent aspects of meaningful behavior is its coherence: our actions may follow each other conditionally, an act may be accomplished in order to be able to perform other acts, actions are performed as executions of intentions which result from processes of motivations and the setting of goals, represented as mental purposes. The same holds for speech acts and speech act sequences (van Dijk, 1980b: 58).

Besides, he also admits (marked cases of) incoherent behaviour:

To be sure, we also are sometimes incoherent, but in many cases this incoherence is the marked case, something which requires specific interpretation, which is or should be noted and which eventually may be sanctioned. (van Dijk, 1980b: 58)

Therefore, rather than discarding traditional linguistic notions, the view taken here is

that they can still be used to distinguish between instances which follow (more or less) conventional expectations of coherence, relevance and cooperation, and those which do not. However, when analysing all these ‘marked’ instances (and before characterizing them as such in video-interaction), rather than just labelling them as ‘incoherent’, ‘non-cooperative’ or ‘irrelevant’, we may admit that the extant linguistic tools are hardly apt for describing them in detail and other approaches and tools are to be attempted.

In other words, in order to fully describe the communicative patterns of video-interaction, a general framework is needed which can explain how meaning is made and sign-making is shaped in *both* coherent and incoherent exchanges, a framework which conceives (successful) communication beyond the interlocutors’ mutual understanding. This general framework needs to keep the pace with and account for contemporary social changes, which are leading to transformations in the ways we communicate and interact, ways in which, it is argued here, principles such as coherence, cooperation and (mutual) relevance are backgrounded in favour (or reshaped in terms) of the pursuing of diversified interests and individual participation, of creative uses and re-uses (i.e., transformations) of other people’s texts, which become semiotic resources to manipulate in the creation of new ones.

The era of late modernity is, by common consent, regarded as a period of fragmentation, of disparateness, of dispersion. We would not expect representational practices to be immune from this phenomenon. In an earlier period, that of seeming monomodality, representation was seen as coherent, as integrated, and as cohesive, as a reflex of social arrangements and practices which were similarly cohesive and stable [...] ‘lines of demarcation’ were clear and were kept clear (and conflicts were precisely about ensuring that the lines, the boundaries, were clear [...]). Mass media production processes were an example of that. The reporter reports, the sub-editor sub-edits, [...] and so on. [...] The age of desktop publishing and website design has blurred such lines of demarcation, and in many cases has already done away with them altogether [...] and new practices for which no scripts as yet exist are coming into being. (Kress and van Leeuwen, 2001: 46)

Kress and van Leeuwen’s theoretical perspective, i.e., social semiotics, seems to be particularly suitable for the analysis of the kind of ‘stuff’ that constitutes video-interaction. The next section introduces its main framework and categories.

### **3 SOCIAL SEMIOTICS**

#### **3.1 Contemporary social (semiotic) changes**

As anticipated, theories are the fruit of their times and, in order to give an adequate account of reality, they need to keep the pace with the changes which occur in society. Semiotic changes – the changes in the ways we make meaning and



communicate – are always the product of larger social, economic and political changes, so that, to be able to describe and explain the former, it is necessary to consider the latter.

Semiotic resources have been produced in the course of social/cultural/political histories – histories which of course keep on going. New social, cultural and political needs lead to new ways of communicating and to new communication technologies – as well as to new communication theories. (Kress and van Leeuwen, 2001: 122)

The contemporary ‘transfer of power from state to market’ (Kress and Pachler, 2007: 7) has led to a change in priorities. If the former favoured homogeneity in its preferred social identity of ‘citizens’, the latter ‘is interested in a high degree of differentiation’ (Kress and Pachler, 2007: 7) of consumers, its preferred social identity. In market-led societies, ‘identity is shaped through consumption, rather than through the achievement of a place in a social structure [...] Agency is exercised as choice from commodities’ (Kress and Pachler, 2007: 7).

The shift from a social organisation around class to a social organisation around lifestyle is, semiotically as well as economically and socially, of the greatest significance [...] The issue is that semiotically speaking the culturally dominating paradigm in the public domain is now that of ‘lifestyle’ [...] Semiotically the shift entails a distinct move towards (greater) individuation: that is, the self-definition of individuals through forms of consumption [...] in which individuation is more intensely emphasised [...] individuation is achieved through consumption of commodities as signs [...] There is pressure on social individuals to differentiate themselves in their individuality through semiotic practices (Kress and van Leeuwen, 2001: 35).

Nowadays, combination, arrangement and *bricolage* of resources (of ‘templates’) are signifiers of differentiation and individuality, and shape the notion of production in new terms:

Although individuals can be made aware of the fact that their choices are also the choices of millions of ‘people like them’, across the globe, they nevertheless *feel* that their style is primarily individual and personal and that they are making creative use of the wide range of semiotic resources made available to them by the culture industries. (van Leeuwen, 2005: 146)

New meanings are assigned to signifiers each time they are used; at the same time, signifiers are more likely produced through selection, recontextualization, transformation and assemblage of pre-existing ‘templates’ (Kress and Adami, 2009). Consequently, contemporary communication is characterized by:

- the (seeming) fracturing of the self into multiple identities as well as the membership of a wide range of user groups and communities of practice;
- the lack of shared cultural experiences as a consequence of a move away from a centrally determined broadcast content and media of transmission and the move towards a ‘distributed’ culture and a model of knowledge assembly;

- the increased fragmentation of mainstream culture into scenes, and the sub-cultures of life styles, each with their own practices;
- the individualisation of social and cultural experiences on the basis of the principles of bricolage.

(from Kress and Pachler, 2007: 14)

What Kress and Pachler are outlining here is the socio-cultural configuration of the current notion of 'social networks',<sup>19</sup> often represented in terms of a newly shaped ruling metaphor of 'participation' and 'production', where structures of power have changed

from "vertical" to "horizontal" and from hierarchical to more *open, participatory relations* [...] While this may be an illusion, seen in wider frames, it is also the case that young people act within such an understanding of power (Kress, 2008).

Horizontal power, production and participation involve changes in the way authority and authorship are conceived:

[w]here previously authorship had been regulated and protected by legal means, in wiki-like production, authorship frequently is not an issue; texts are open to constant modification [...] In *downloading, 'mixing', cutting and pasting, 'sampling', re-contextualization*, questions such as 'where did this come from?', 'who is the original/originating author?' seem not an issue [...] the very same processes are in use at 'higher levels', as 'pastiche' in Post Modern art forms (Kress, 2008).

Rather than a certificate of ownership, authoring is conceived as (individual and collective) action, which constitutes newly shaped and newly conceived 'forms of sociality', which constitute 'inceptive communities' (Kress, 2008). Bonds among members are as loose as these forms of communication. Authoring-as-action shapes practices and affects rules and structures.

Video-interaction is one of the expressions of social networking, so, in order to analyse it adequately, we need a model of communication and interaction which can account for multifaceted, fragmentary and individualized practices which are driven by diverse micro-interests and goals and which, only secondarily (and incidentally), may achieve a common and more global goal or aim. In a world where '[c]ommunication [is] the social glue holding the show together' (Kress, 2008), a new social role is fostered, i.e., the 'rhetor':

[i]n a world of meaning (the semiotic world) and values (the ethical world) marked by instability and provisionality, every event of communication is in principle unpredictable in its form, structure and 'unfolding'. The absence of frames requires of each interactant on each occasion an assessment of the social relations which obtain and of the resources available for shaping the encounter. This demands a

---

<sup>19</sup> The notion of 'social network' is widely used; cf., among others, Wellman (1996), Wellman and Berkowitz (1988), Scott (1991), Wasserman (1994); for a recent review, cf. Freeman (2006).

rhetorical approach to communication rather than the prior approach of *competent performance*, where inwardness with ‘grooved convention’ was sufficient for competent communication (Kress, 2008).

In the ‘communication required for full participation in the new media world’,

[t]he *rhetor* has to make an assessment of all aspects of the communicational situation: her/his interests, the characteristics of the audience, the needs of the issue at stake, and the resources available for making a representation and for its dissemination. (Kress, 2008)

When authorship is conceived as action which defines membership, participation is made of rhetorical acts. Participants in video-interaction, as will be seen, are rhetors of their age and of that specific context; they produce their semiotic acts by exploiting every possibility (medium, resources, contexts, previous texts etc.) to their own purposes, thus transforming conventions even before these are established (cf. the analysis in Chapter 4).

### 3.2 Relevant concepts and categories

Videos are multimodal texts which make meaning by employing a wide range of signs made through different semiotic modes (such as gestures, facial expressions, music, colour effects, drawing and images, etc.), deployed both in time and space, together with the specific resources of dynamic images (such as camera angles and filmic cuts, for example). Theories such as the ones discussed earlier, which have devised their categories and tools of analysis concentrating mainly (if not even exclusively) on verbal (and sometimes only written) language, are not able to handle these dynamic multimodal texts. Indeed, many of the videos of the corpus do not even deploy verbal language at all and a great effort would be needed to adapt linguistic categories so as to describe a wink or a camera angle or a soundtrack, for example, not to mention the meaning which is made by the co-deployment of these resources.

Stemming from Hallidayan linguistics (1978), social semiotics (Hodge and Kress, 1988) has adapted its categories, extending them so as to describe the meaning which is made through other semiotic modes, so that it is particularly apt both to the object of study of the present work (i.e., multimodal texts) and to its pragmatic aim (i.e., to be read by linguists; it indeed ‘speaks’ their language, while adapting it).

Social semiotics conceives signs as ‘semiotic resources’ (van Leeuwen, 2005: 3). Rather than arbitrarily determined by conventional rules fixed into a code, the association of signifier and signified into a sign is socially, culturally and situationally motivated and is the result of the sign-maker’s interest and of her available resources at the time of sign-making. ‘Signs are always motivated by the producer’s “interest”, and by characteristics of the object’ (Kress, 1993: 173).

Sign-makers thus ‘have’ a meaning, the signified, which they wish to express, and then express it through the semiotic mode(s) that make(s) available the subjectively felt, most plausible, most apt form, as the signifier. This means that in social semiotics the sign is not the pre-existing conjunction of a signifier and a signified, a ready-made sign to be recognized, chosen and used as it is, in the way that signs are usually thought to be ‘available for use’ in ‘semiology’. Rather we focus on the process of sign-making, in which the signifier (the form) and the signified (the meaning) are relatively independent of each other until they are brought together by the sign-maker in a newly made sign [...] In our view signs are never arbitrary, and ‘motivation’ should be formulated in relation to the sign-maker and the context in which the sign is produced, and not in isolation from the act of producing analogies and classifications. Sign-makers use the forms they consider apt for the expression of their meaning, in any medium in which they can make signs. (Kress and van Leeuwen, 1996, 2006: 7-8)

Hence signifiers are semiotic resources which are newly made into signs (associated to a signified) every time they are used; this means conceiving them as having a ‘meaning potential’, which is differently actualized whenever a given resource is used by a sign-maker:

So in social semiotics resources are signifiers, observable actions and objects that have been drawn into the domain of social communication and that have a theoretical semiotic potential constituted by those past uses that are known to and considered relevant by the users of the resource, and by such potential uses as might be uncovered by the users on the basis of their specific needs and interests. (van Leeuwen, 2005: 4)

Within the social semiotic framework, multimodal analysis (Kress and van Leeuwen, 1996, 2006; Kress and van Leeuwen, 2001) has developed a set of tools and a methodology which interpret the underlying ‘grammar’ according to whose patterns meaning is made through the co-deployment of signs made in a variety of semiotic modes. Multimodal analysis relies on the three metafunctions originally proposed for language by Halliday (1978) and adapted to multimodal representations in Kress and van Leeuwen (1996, 2006), namely:

- the ‘ideational’ function, i.e., the function of conveying representations of the world, in terms of processes, participants and circumstances (in Halliday’s terms ‘what the text is about’);
- the ‘interpersonal’ function, i.e., the function of enacting the relations between interactants, their attitudes and social roles (‘who produces the text for whom’);
- the ‘textual’ function, i.e., the function of orienting the representation and relating its constituting elements to one another and to the whole text (‘how the ideational and interpersonal are organized in the text’)<sup>20</sup>.

---

<sup>20</sup> The three meta-functions are used here as descriptive tools in Chapters 5 and 6.

Following Halliday (1978), also grammar is conceived as a social resource<sup>21</sup>, rather than a fixed code separated from its actual sign-makers' interested use. In this view, traditional distinctions between semantics and pragmatics, between *langue* and *parole* (Saussure, 1931), between structure and use are irrelevant in social semiotic multimodal analysis:

signs may not be divorced from the concrete forms of social intercourse (seeing that the sign is part of organized social intercourse, and cannot exist, as such, outside it). (Hodge and Kress, 1988: 18)

Therefore, rather than semantic meaning vs. pragmatic meaning-in-use, social semiotics talks of meaning potential vs. meaning which is historically, culturally, socially, situationally and individually actualized. Rather than merely *used*, signs are always newly made, every time a sign is needed. And sign-making and meaning-making occur both in what is traditionally called text production ('articulation' in multimodal terms, Kress and van Leeuwen, 2001: 8) and in text reception ('interpretation', 2001: 8):

it is essential to stress that production is common to both articulation and to interpretation. The general principles of semiosis, of sign-making, are the same in both cases. (2001: 41)

Multimodal Discourse (Kress and van Leeuwen, 2001) conceives communication as always occurring when there is interpretation, no matter if what is interpreted corresponds to the meaning intended by the producer of the text (unlike the above discussed models of communication). In fact, the opposite is rather the normal unexceptional process, in that there is never a situation where what is interpreted corresponds exactly to the meaning intended by the producer of the text.

We define communication as only having taken place when there has been both articulation and interpretation. (In fact we might go one step further and say that communication depends on some 'interpretive community' having decided that some aspect of the world has been articulated in order to be interpreted). (2001: 8)

Conventions are established (as well as constantly negotiated and transformed) as the fruit of (often conflicting) power relations among sign-makers: 'rules, whether written or unwritten, are made by people, and can therefore be changed by people [...] To be able to change rules you need power' (van Leeuwen, 2005: 47).

[R]ules can never control every detail of what we do. In a sense every instance of sign production and interpretation is new. We never just mechanically apply rules. Every instance is different and requires adaptation to the circumstances at hand. (2005: 50)

Furthermore, '[i]n the case of new semiotic resources and new uses of existing

---

<sup>21</sup> Cf. Halliday's notion of grammar as a 'resource for making meanings' (1978: 192) rather than a pre-given code.

semiotic resources, old rules are cast aside' (2005: 50).

The perspective outlined by social semiotics is particularly apt to frame the analysis of video-interaction, since the latter is a new type of communication, in which conventions are not already fixed and established, but are rather being constantly negotiated among participants and with the Website owners. Even more, in contemporary forms of communication such as video-interaction, rather than rules, one can observe regularities (and variations) in use, which are constantly modified and transformed.

Apart from the novelty of video-interaction-as-process, also the texts of video-interaction (online videos) are new. Indeed, although we are ordinarily exposed to dynamic images distributed by mass media, only recently 'ordinary' people have started to make their own videos and share (i.e., broadcast) them online. Filmic conventions are shaped in different ways in online videos, so that we can only partially rely on them for the interpretation of the so-called user-generated videos<sup>22</sup>. Furthermore, video-interaction can be confidently considered as the first type of communication in which a video can 'reply' to another video. As a consequence, at any given time, what is possible to do in video-interaction and what is done are the result of various practices which respond to different (and sometimes conflicting) interests, rather than of established conventions described (and consequently prescribed) in film-making theories and textbooks.

In reason of this, a perspective such as the one of social semiotics, which conceives communication as a socially grounded diversified process of perpetual transformation, in which rules are transformed together with signs, in which use is a constant (re)making of sign, is more suitable to describe a new and 'loose' form of communication such as video-interaction.

The pre-theoretical richness of communication keeps getting lost in definitions of communication based on information-theoretic, mentalistic or linguistic models, which place their core, respectively, in 'exchanges of information', in 'inferences of mental states' or in 'language mediation'. Each excludes what the other insists is essential, and none captures either the consequentiality or the intimacy that can be gained in human engagements. (Costall and Leudar, 2007)

---

<sup>22</sup> Film Theory (Eisenstein, 1949; Eisenstein, 1994; for a recent review and introduction on Film Theory, cf. Stam, 2000) has extensively investigated the meaning made by means of dynamic images in cinema. Yet, on *YouTube*, videos are made by 'everyday' people, so that traditional filmic conventions intertwine with new ones, according to the available resources and to the new conventions which are being negotiated and established in this semiotic space. Throughout the work the term 'amateur' is avoided whenever possible, since the distinction between amateur and professional video-maker is now hard to define, especially because of the wide popularity of video-sharing Websites such as *YouTube*. Moreover, this distinction would be useless for the aims of the present work, since the only possible distinction between the two categories would be in economic terms (i.e., whether the video-maker earns a living out of making videos), which however would not say anything on the sign-making and conventions on *YouTube*.

Costall and Leudar are by no means social semioticians and, in their article, the above quoted statement is made as a criticism to Theory of Mind; however this excerpt is a useful resource which can be interestingly recontextualized (and possibly misinterpreted) here, so as to give a further clue of the ‘global coherence’ of the topics discussed so far.

The next two sections introduce the notion of ‘affordances’ and that of ‘interest’ and their derived heuristics which are used in Chapter 4 and in Chapters 5 and 6 respectively, in the analysis of the process of video-interaction (Chapter 4), and of its texts (Chapters 5 and 6).

### 3.3 The notion of affordances

Every form of communication is mediated; to represent we use either ‘embodied’ or ‘disembodied’ resources (Norris, 2004), i.e., we use a technology – our body in the case of embodied modes, and/or objects, in the case of disembodied modes – which has its own materiality and social function, in terms of what it materially enables and socially allows us to do with it. In social semiotic terms, every medium (and semiotic mode) that we use to communicate and to represent has ‘affordances’ (Kress and van Leeuwen, 1996, 2006), in terms of both material and social possibilities and constraints, i.e., in terms of what it is materially (im)possible and socially (non)permitted to do with it. The view taken here is that sign-makers make an interested use of these affordances, generating diverse semiotic practices which, when made socially powerful, can give rise to a change in the affordances themselves. In sum, practices bring forth structural changes, i.e., in Saussurian terms, the acts of *parole*, acknowledged and made ruling, influence the *langue*.

In contemporary forms of communication, especially in the social networking ones such as video-interaction – characterized by the horizontal, loose and fragmented features previously discussed – this active influence of semiosis (i.e., semiotic processes) over affordances and ‘grammar’ is particularly manifest.

Online video sharing employs a wide range of technologies for the production and the distribution of texts. To make a video you can use a video-camera or a Webcamera or you can simply use the computer and make a video out of a slideshow of pictures. You may have differently available a range of software tools for video making and editing and you may master them at various extents. The viewing of other videos can prompt you to use new technologies and resources (how-to videos are a highly populated genre on *YouTube*).

Uploading videos on *YouTube* and distributing them needs to take account of what is allowed and permitted to do on the interface by the Website owners, i.e., what the affordances of the Website are. Responding to a video with another video implies exploiting the video response option affordances – in terms of what is technically

possible and socially accepted on the Website – for given purposes, which may differ among participants.

The tension between what the structure affords materially and socially and the practices (i.e., the effective uses of these affordances by the participants according to their interests) is constant in all forms of communication in general, and in video-interaction in particular, given that it is a new and thus relatively unconstrained genre. Even more, new and unexpected practices may give rise to changes in the affordances, in terms of implementations made on the interface so as to favour (and thus institutionalize) some of these practices and hinder others.

This theoretical assumption constitutes the framework for the first chapter of analysis of video-interaction, which investigates the rise of the phenomenon and its extant development in terms of affordances and practices, and of their mutual intertwining and influence.

### **3.4 The notion of interest**

The interest of sign-makers, at the moment of making the sign, leads them to choose an aspect or bundle of aspects of the object to be represented as being criterial, at that moment, for representing what they want to represent, and then choose the most plausible, the most apt form for its representation. (Kress and van Leeuwen, 1996, 2006: 13)

In a social semiotic perspective, representation is a matter of interested choice, both in relation to the characteristics of the world to be represented and in relation to the forms used to represent it. This choice is made by the sign-maker according to her interests at that given time, among the semiotic resources which are available to her at that given time. The meaning-maker (i.e., the one who interprets the sign) also selects the signifiers of the text which are salient to her and associates them to signifieds according to her interest at the time of interpreting the text and on the basis of the semiotic knowledge available to her at that time. It is understood that semiotic resources are not just ‘signs’ conceived as forms associated to contents in the mind of the sign-maker; rather, they are forms taken from and contents interpreted from all the existing texts experienced by the sign-maker.

In interactional exchanges (e.g., a video responding to another), this ‘interested’ sign-making is a cyclic process:

- a text (e.g., a video) is made through an interested selection of form (among the available resources) and content (i.e., the criterial aspects of the world to be represented);
- its interpretation by the viewer (i.e., a meaning-maker) is driven by an interested selection of the criterial aspects of the interlocutor’s representation;
- in turn, this interest-driven interpretation shapes the criterial aspects and the



most apt form chosen to ‘respond’ to that text.

In other words, a text prompts a set of possible (interpretations and) responses. The actual response to the prompt, i.e., which of the prompted possibilities is taken up in the response, depends on (a) the sign-maker’s (social, cultural, psychological and situational)<sup>23</sup> interest and (b) on the resources available to her at that time. Hence exploring sign-making in a chain of semiosis implies analysing how a relation of prompt-response is actualized in the texts of the interaction, which can give clues of the various sign-makers’ interests.

In other words, the heuristic assumption that drives the second and third chapters of data analysis (the ones devoted to the analysis of the texts) is that the interested actualization of semiotic resources into signs has its interactional counterpart in an interested actualization of prompts into responses. If the former is the focus of the analysis of individual texts, the latter is the focus of the analysis of interactional exchanges. In the present work, this means exploring how this interest-driven prompt-response relation is actualized in video-interaction.

## **4 CATEGORIES USED IN THE ANALYSIS**

### **4.1 Video-interaction as process and video-interaction as texts**

A preliminary distinction needs to be done between the process of video-interaction and its products. Clearly, the two cannot be separated, both as phenomena and as object of study; indeed, by definition, while the former produces the latter, any investigation of the former can only be done through the observation of the latter (i.e., the texts which are viewable on the Website as a result of the interaction-as-process). The categorization of reality into processes and products is of course merely analytical and descriptive. At the level of processes, sign-making is conceived in terms of semiotic acts; at the level of products, it is conceived in terms of texts. In the case of embodied resources the two coincide in time, so, a gesture is an act and a text at once. Since video-interaction is asynchronous and its products are disembodied texts (i.e., video clips linked one to another which can be re-played at will without the presence of their producer), for the pragmatic purposes of the analytical description, the process of video-interaction (i.e., the affordances and the practices) constitutes the main focus of the first chapter of analysis (Chapter 4),

---

<sup>23</sup> The interest ‘is a complex one, arising out of the cultural, social and psychological history of the sign-maker, and focused by the specific context in which the sign-maker produces the sign.’ (Kress and van Leeuwen, 1996, 2006: 7). In Kress (2009) ‘interest’ is defined as:

the ‘condensation’ at the moment of representation of an individual’s (social) history; a sense of who they are in the social environment of communication as well as a sense of the salient features of the environment in which the prompt occurred. These lead to the selection of that aspect of the phenomenon treated as ‘criterial’ for the purposes of representation.

while the products (i.e., the videos and their textual relations) are investigated in greater detail in the second and third chapters of the analysis (Chapters 5 and 6).

Here, in the next two sections, categories and terminology are introduced which refer to video-interaction as process and video-interaction as texts respectively.

#### **4.1.1 The prompt-response relation in the process**

A video is a complex sign. The view of a video ‘provokes’ in the viewer various types of responses. In general terms, a response is anything which is prompted by something (cf. Merriam-Webster dictionary definition of ‘response’: ‘something constituting a reply or a reaction’<sup>24</sup>).

When made by an animate entity, a response defines the prompt as a meaningful act or fact; whenever there is a response there has been meaning-making. Even when the response is prompted by an event which is not meant to communicate, the response makes its prompt meaningful. Thus, when an actor (an animate entity) perceives an event and reacts to it, the actor has made some meaning out of this event; for example, when an animal runs away from a fire, that animal has made some meaning out of the fire (has understood its potential danger). However, neither the event nor the animal’s reaction are semiotic acts in themselves, in that they are not made of representational signs so as to communicate meaning (although they can be interpreted as such by human beings).

Conversely, video-interaction – as all forms of human interaction – is made of semiotic acts created and interpreted by meaning-makers in interaction. Online videos are complex signs (i.e., texts) made by human beings with the main purpose of communicating. Videos are texts produced by a sign-maker, i.e., its author<sup>25</sup>. They are replayed (watched) by another meaning-maker, the viewer.

The viewer’s response to a video can be a semiotic act (a meaningful act, i.e., an act which generates a text) of a different type from the prompting one. So, on the Website, the viewer can respond to the viewing of a video by posting a written comment, by rating the video, by bookmarking it as favourite, by flagging it as inappropriate, by subscribing to its uploader’s channel, and so on. These are all acts which respond to the prompt and communicate something in their turn; like all semiotic acts, they potentially prompt something in their turn and they may well aim at it; e.g., comments can aim at receiving a feedback from the author of the video (or open a debate about the video), rating can contribute to promote or debase the video, bookmarking aims at having the video readily available for future browsing,

---

<sup>24</sup> <http://www.merriam-webster.com/dictionary/response> (Retrieved 26 October 2008).

<sup>25</sup> For the sake of simplicity, the video maker and the video uploader are treated as one participant, yet they may well be different persons (or even groups of people), as in the case of somebody making a video and asking somebody else, who has a *YouTube* profile, to upload it (cf. Section 4.2).

inappropriate flagging can aim at banning the (You)Tuber from the community, subscribing can aim at becoming ‘friend’ with the video uploader and get a reciprocal subscription, and so on. In a similar way, outside the Website, the viewing of a video can prompt the sending of the video-link to email contacts to make it known (and, eventually, ‘viral’), or the insertion of the link into one’s own blog or homepage; one can write an article or a review about it or can talk about it in face-to-face encounters, etc. One can even (negatively) respond to the viewing of a video by deciding not to go to the Website again or not to ever watch that video again.

In all these cases, the interaction that takes place between the two meaning-makers (i.e., the video uploader and the viewer) is what I term here *heterogeneous*, in that the video has prompted its viewer to respond with everything *but* a video. The response belongs to a different genre (i.e. type of text<sup>26</sup>) than the prompt and it is realized in different modes and through different media from the ones which prompted it.

In turn, when a video prompts the viewer to make a video in her turn, the prompt-response relation is *homogenous*, in that it involves the same type of complex signs. Furthermore, when a video prompts the viewer to make a video and post it as a video response to the first one, the interaction that takes place is an homogeneous *exchange* between the two participants; it is made by two complex signs of the same type (i.e., videos), one of which addresses the other as a response to it.

Therefore a first distinction in the prompt-response relation is between the ones who, after watching a video, make a video response, and the ones who do not (and do something else). As understood here, video-interaction, concerns the former type of prompt-response. Hence, *video-interaction is a process of exchange of (at least) two homogeneous texts (i.e., videos), one of which is a response to another.*

Given that it is the existence of the video response that defines the exchange (i.e., a prompt can be said to be realized only when it gets a response), *the unit of analysis of video-interaction is composed of (at least) two videos (and the link between them): the prompting one and its video response. This is the basic syntagm of video-interaction.*

Regardless the content and the modes through which the content is realized, a first prompt-response relation takes place at the level of semiotic acts, i.e., in the process of video-interacting: the uploading of a video prompts the uploading of a video response. This constitutes a basic video-interactional exchange.

---

<sup>26</sup> The labels ‘genre’ and ‘text type’ are variously used in the literature of Genre Analysis (Beghtol, 2001; Bhatia, 1993; Biber, 1989; Biber and Finegan, 1986, 1994; Chandler, 1997; Crowston and Williams, 2000; Crystal and Davy, 1969; Fairclough, 1992; Görlach, 2004; Halliday and Hasan, 1985; Hoey, 1983; Paltridge, 1995, 1996, 1997; Sampson, 1997; Steen, 1999; Swales, 1990; Taylor, 1989; van Dijk, 1977); often they are used interchangeably. For a discussion, cf. Lee (2001).

Video-interaction is the process of interacting *by means of* videos. As a matter of fact, it is a process which is constituted by various semiotic acts:

1. the uploading of a video-file<sub>1</sub> by (You)Tuber<sub>1</sub>;
2. its viewing by (You)Tuber<sub>2</sub>;
3. the (making of a video and its)<sup>27</sup> attempted link by (You)Tuber<sub>2</sub> as a video response to video-file<sub>1</sub>;
4. the viewing of video-file<sub>2</sub> by (You)Tuber<sub>1</sub>;
5. the acceptance or denial of video-file<sub>2</sub> as a response to video-file<sub>1</sub> by (You)Tuber<sub>1</sub>.

The semiotic act in 3. is a response to what prompted in 2. (to the viewing of video-file<sub>1</sub>); the viewing in 4. is a response to what prompted in 3.; the semiotic act in 5. (either the denial or the acceptance) is a response to what prompted by the viewing in 4. The denial in 5. results in unsuccessful video-interaction, which is unobservable on the Website (no trace of the denied attempt is visible), so that the two videos remain unrelated texts on *YouTube*. In turn, the acceptance in 5. realizes a successful instance of video-interaction and results in a communication exchange between two participants by means of two videos, one of which is the response to the other.

The successful process of video-interaction is observable on the Website through its product, which is composed of three elements: (a) video<sub>1</sub>, (b) video<sub>2</sub>, and (c) the link (visible on the Website, on both video<sub>1</sub> and video<sub>2</sub> pages) which establishes video<sub>2</sub> as response to video<sub>1</sub>, its prompt. The three elements taken together are the minimal unit of analysis of video-interaction, which is then an exchange constituted by a prompting video and its response.

From the above, it becomes evident the reason why theories which define successful instances of communication in terms of correct interpretations of the communicator's intended meaning cannot account for video-interaction. Indeed in video-interaction, successful exchanges are determined by the initial (You)Tuber's acceptance of the responding (You)Tuber's video as a response. This acceptance is not bound to the participants' interpretation of each other's intended meaning but is a consequence of their varied (and sometimes conflicting) interests, as evidenced in the analysis.

#### **4.1.2 The prompt-response relation in the texts**

As a text, a video is a complex sign composed by a series of signs made in a variety of semiotic modes, both in simultaneity (through the spatial arrangement of visual elements combined with sound and speech) and in sequence (through the succession in time of visual and auditory elements). Any of these elements and any combination of them can be perceived as salient and can be taken up – i.e., corresponded to – in

---

<sup>27</sup> A video response can be made also by linking a previously existing video; hence the making of the video is an optional semiotic act (cf. Chapter 4, Section 2.3).

the interaction. Therefore, a prompt-response relation does not exist only at the level of semiotic *acts*, as discussed above, but also at the level of *texts*, in what is represented in videos and, specifically, in all the elements represented in the prompting video which are variously taken up in the response.

As with the semiotic acts, also with the texts, a prompt can be said to be realized only when it is responded. Therefore also here the prompt-response relation between the two videos is actualized and defined by the existence of a response. Consequently, *the unit of analysis of the interaction between videos is constituted by any element in the prompting video and its corresponding element in the response.*

This unit of analysis is less objectively identifiable than the one of the video-interaction exchange (i.e., of the process). Indeed, if the unit of analysis of the process is instantiated by a visible element on the interface, i.e., the video response link, the unit of analysis between videos can be (subjectively) identified only by the viewer when watching the two videos. Yet it is the existence of this prompt-response relation between videos which can tell to what extent a video is *related* to the one it responds to. Indeed, only if the viewer can establish a relation between what is represented in the response and what is represented in the responded video, she will consider the response as related to the first (and hence, to some extent, coherent). Otherwise, she will consider the two videos as unrelated, in spite of the existence of a visible response link between them on the Website, and will consider this link to be used for other purposes than the specific interaction in the exchange, i.e., what is traditionally considered as ‘spam’<sup>28</sup>. Furthermore it is only the analysis of the prompt-response relation between texts that can tell something of *how* participants interact by means of videos.

Therefore, by investigating the elements in videos that are involved in a prompt-response relation, one can tell:

1. how participants make meaning in interaction through the video response option;
2. how texts prompt them to respond in theirs;
3. how they relate their texts to others in interaction.

#### **4.1.3 The analyses of texts and processes: responded prompts**

Following Heritage (1984), the work of Hindmarsh and Pilnick stems from the assumption that ‘every action in interaction is both context shaped and context

---

<sup>28</sup> When referred to video-interaction, the term ‘spam’ is always between inverted commas, since it is deemed inappropriate for video responses which are actually displayed as such to a given video. Indeed, precisely because the thread initiator has the power to filter responses out, when a video is shown on the interface as a response to another one, it can reasonably be considered as not ‘disturbing’, unlike usual spam. Yet, this position is not shared by Benevenuto et al. (2008b).

renewing, both organized in the light of the prior action and framing the next' (2007: 1401), so that 'conduct is both retrospectively and prospectively organized' (2007: 1401). Although they analyse real-time organized interaction of teamwork (in anaesthesia rooms), this general assumption can be kept valid also for asynchronous types of non-specifically goal-oriented interactions like video-interaction.

Restating Hindmarsh and Pilnick in the terms used here, every semiotic act does not only respond retrospectively to a prompt, but is also prospectively organized, i.e., it is made so as to prompt a response in its turn. It is intended to have communicative effects. Analogously, in video-interaction, any video response can be intended to prompt various responses, either heterogeneous (like comments, views or subscriptions, as discussed above), or homogeneous, so that a response can be responded in its turn, thus giving the way to a further exchange, in terms of a sub-level in the thread. Furthermore if the first video prospects a follow-up, the response can be modulated so as to be included in it, and, in its turn, the follow-up will be a response to it by the first participant in the exchange. In this case the structure of the exchange will be composed of (a) an initiating move, which prompts (b) its responses, which prompt (c) the follow-up made by the first participant.

However, if what has been prompted is observable in the response, what the response can prospectively (or is intended to) prompt is hardly observable in the response itself. Only what is effectively taken up by a subsequent response (or by the follow-up) can be said to have a realized prompting effect.

Therefore, as in the process of video-interaction only successful instances can be observed, so that the minimum unit of analysis is constituted by a prompting video and its response, also for the prompt-response relation within texts, only its positive realizations can be observed, so that the minimum unit of analysis is constituted by a prompting element in the responded video and its corresponding element in the response. In other words, the intentions of the interactants cannot be determined; rather, their interests can be inferred by the types of prompts which are selected from the initial video and actualized in its responses.

Besides, by analysing how the prompt-response relation is variously realized in the thread (i.e., by comparing the different prompts that are taken up in the various responses), one can map the (more or less) wide range of meaning potentials of the first video which are *actualized* in the responses. Within this range, there may well be the intended (prospectively organized) prompts along with the unintended ones (but retrospectively accepted by the first (You)Tuber, in her approval of the response).

#### **4.1.4 Caveat: the prompt-response relation in a social semiotic perspective**

As discussed above, the label 'response' is here used in two cases:

1. in the case of processes, to name the actualized relation of the interactional exchange, taking up the label given by the interface to the link that (You)Tubers can create among videos (the ‘video response’ option);
2. in the case of texts, to identify one side of the relation of the elements occurring between videos as texts, i.e., the elements of the initial video which are taken up by the video responses (the other side being the prompt).

In both cases, the term ‘response’ is used here totally free from any mechanical meaning. As described earlier, it is used with reference to its primary meaning of ‘a reply or a reaction’ to a prompt, and no implication is ever made to the deterministic and mechanistic discourses that may be implied by using the term in association with physical stimulus-response relations. Indeed, as anticipated earlier, in a social semiotic perspective what is perceived as prompt always depends on the interests of the meaning-maker (the viewer of a video), and how a meaning-maker responds to a prompt also depends on her interests. These interests are determined by social, cultural and individual dimensions which characterize the meaning-maker’s perspective at the time of her meaning-making and by the specific situation in which the communication takes place, as well as the material resources available to her in that time and in that situation.

In other words, there is nothing automatic, deterministic or mechanical in the prompt-response relation, to the extent that, as said, what a video *potentially* prompts cannot be predetermined but rather inferred subjectively. It is the fruit of the subjective meaning made by the researcher as a socially, culturally and individually specified viewer, according to her interests at that given time.

Conversely, what a video *actually* prompts is observable on the basis of what is responded, in terms of representations, representational resources, topics (semantic fields), and their specific organizations in the text. This is the reason why the analysis is empirically carried out on existing instances of video-interaction.

What is taken up as a prompt by a response can be considered totally irrelevant by another. So, the response determines what is prompted and the type of response is unpredictable, because it depends on the respondent’s specific making of meaning, i.e., on her interests and purposes, which are influenced by her social role and position, her cultural background, as well as her individual interests and purposes at that time in relation to that communicative situation.

Yet, in spite of these many variables, patterns of regularities and variations in the prompt-response relation are clearly perceivable in the actual threads of video-interaction (cf. Chapters 5 and 6). Furthermore, what the response takes up from its prompting video determines (or contributes to determine) the extent to which it is related to it. In other terms, how the prompt-response relation is configured determines the extent of coherence and cohesion of the exchange. This is the reason why investigating this socially, culturally and situationally specific prompt-response

relation is deemed useful. It may be extended to other types of human interaction and may reveal itself as a useful analytic tool to investigate how we relate texts one to another, how we perceive coherence and how we make meaning in interaction.

## 4.2 The participants in video-interaction

The participants in a video-interaction exchange are to be identified in the two profiles of the uploader of the prompting video and the uploader of the video response. One will be termed here as initial (or responded) (You)Tuber, the other as responding (You)Tuber (or respondent). Each (You)Tuber is generally understood as being the author of the video, yet the author may be different and may also be credited in the video or in its paratext.

By author of the video it is meant the person(s) who creates the video-file, not the person(s) who creates its content. So, for example, a video which is an excerpt of a TV-show is authored by the one who has selected the excerpt, turned it into a video-file and then uploaded it on the Website. The creator of the TV-show is not at all to be considered the author of the video, since the video on *YouTube* is a different text in a different context (i.e., made of a different material – i.e., a *Flash* file – distributed on a different medium, by a different person, to a different audience, for different purposes), so that the TV-show functions like a ‘direct quote’ in the video-file. Analogously, when somebody emails a news article to somebody else, the author of the email is not the author of the article and anyone who responds to the email responds to its author, not to the journalist. In other terms and in spite of the ongoing debate on the notion of ‘authorship’, the view taken here is the one introduced earlier of ‘authoring-as-action’ (cf. Section 3.1). In acting semiotically, participants produce texts (also, and always more often, by selecting, copying and pasting, making collages of other people’s texts) and, in producing these texts, they act in a semiotic space which refers to them as the authors.

The person behind the username is not identifiable, nor is it always possible to ascertain whether usernames are used collectively or individually, nor even whether the videos are created and uploaded by the same person or by different ones. Yet this is not the focus of the investigation. The focus here is on the texts, on how these texts make meaning in interaction and on how interactants develop semiotic practices by means of them. In other terms, the focus here is on the ‘semiotic space’ (Gee, 2005) rather than on the ‘community’<sup>29</sup>. It is on the semiotic practices rather than on

---

<sup>29</sup> The long-standing debate on the notion of ‘community’ makes this label potentially ambiguous if not circumscribed to a given theoretical tradition within the academic discourse. The notion of ‘community’ is generally avoided here, favouring in its stead the notion of ‘semiotic space’ (Gee, 2005). When the term ‘community’ is used here, it occurs always in reference to ‘discourse(s)’ – conceived as ‘socially situated forms of knowledge about (aspects of) reality’ (Kress and van Leeuwen, 2001: 4) – to refer to a broad ‘tubing-as-tie’ concept as is ‘talked about’ in videos, in Foucault’s terms (1971). In other words, the related analysis is not interested in ascertaining *what* is that may bind (You)Tubers among them, but rather *what is discursively presented* in their semiotic



members. It is on roles and relationships (even identities) that are communicated, negotiated and constructed in this semiotic space, not on who these people are (admitted that it could ever be possible to say anything about that); cf. also the discussion in Chapter 3, Sections 4.3 and 5.2.

Certainly, the more information available on a (You)Tuber, the more it is possible to infer what influenced a certain semiotic act. The amount of information available online about the (You)Tubers' semiotic production is impressive; this can include their channel profile on *YouTube*, all their uploaded videos, their links to others, their other online pages linked on *YouTube* (*Myspace* and or *Facebook* profiles, Weblogs or Webpages etc.), their subscribers, their 'video-favourites' etc. When deemed useful, all this huge bulk of (self)representations is considered in the analysis. However, this all will always be considered as (further) texts. Ultimately, the analysis here endorses the assumption that if identities presented and represented online are made of texts, so are the ones offline. A little provocatively, if, as assumed, there is communication only when there is interpretation, analogously, identities are (defined) only when someone *interprets* them as such.

### 4.3 The protagonists of video-interaction

Videos can – and often do – have protagonists, as characters represented. The main character represented in the text may represent the (You)Tuber (the equivalent of the pronoun 'I' in verbal texts), and usually does, if not stated otherwise. This means that in videos the '*sujets d'enoncé*' can refer to the '*sujets d'enonciation*' (Benveniste, 1971); in semiotic terms, the represented participants can represent the interactants.

The overall majority of the most responded videos of all time is made of user-generated content and most of them feature the (You)Tuber as the main character. In most videos used as data here, the represented participant is (understood as) the (You)Tuber's persona, who hence refers to the (You)Tuber. When the represented participant is the (You)Tuber's persona, the ideational meaning clearly has direct interpersonal meaning (the persona communicates something about the (You)Tuber-interactant). Indeed, it could be said that when the text is viewed, the represented participant communicates *on behalf of* the interactant.

Therefore, in the following analysis, the term '(You)Tuber' will be used to define the interactant (*sujet d'enonciation*) as communicated by her persona (*sujet d'enoncé*), which can be variously represented. Indeed videos enable a mix of embodied and disembodied signs, so that 'I' can be a face, a gesture, an avatar, a spoken word, a written word, or a logo. These are all signifiers of the (You)Tuber.

---

representations. For a review of the debate on the notions of '(virtual) community', 'community of practice', and 'communities as social networks', cf., among others, Barton and Tusting (2005), Baym (Baym, 2000), Bell (2001), boyd (2006), Jones (1998), Licklinder and Taylor (1968), Slevin (Slevin, 2000), Wellman (1996), Wellman and Gulia (1999), and Wenger (1998).

#### 4.4 The texts of video-interaction

The texts of video-interaction are the videos linked through a response relation. They have paratext (the video homepage on the Website), which makes meaning together with the text. Some of their paratext is given by their (You)Tuber, some is given by the interface (which also determines the whole layout of the video page); some is the product of interaction (responses thumbnails and text comments), some is a mixed contribution of user-generated content and interface (thumbnails of related videos).

Videos allow for (or give a pretension of) spatial and temporal displacement of their participants. Indeed, video-interaction is made by means of asynchronous and disembodied texts. Asynchronicity and disembodiment imply that the text can be replayed at will while its producer is not there.

Like all other texts, videos are representations. Like all other signs, the signs used to represent the participants are selected according to the interest of its producer, and the meaning which is made from interpreting these signs is according to the interest of its viewer. Yet, differently from other complex signs, videos are disembodied texts which make meaning through visual and auditory signs organized in space and time. They are selective and edited visual and auditory ‘capturing’ and ‘recording’ of entities and events. Videos allow for embodied (and disembodied) signs to be enacted in disembodied texts; in this sense, ‘videoblogs’ (videos where the videoblogger appears typically facing the camera and talking to viewers) are a representational step forward of the letter-type, in terms of ‘realism’; they indeed combine features of the letter (disembodiment) and of real-time co-present communication (embodiment).

Of all semiotic modes afforded by face-to-face interaction, only body-contact is excluded in video-interaction, since interactants are not in spatial co-presence. In turn, compared to face-to-face, videoblog defies space and time in that by means of a video an interactant can communicate through embodied modes (but body contact) to anyone, anywhere and whenever the video is played. So, although the ‘author is dead’ (Barthes, 1977a), the ‘reality’ modality of a videoblog lets the viewer perceive the (You)Tuber’s persona in a video *as if* the (You)Tuber herself.

##### 4.4.1 Texts within texts

Respondents can quote by embedding a selection of shots of the original video and can ‘mis-quote’ or elaborate the quote (similarly to reported speech or indirect quotation) by editing the original video, remixing or enacting a selection of formal elements of the original video through their own material. Mis-quotations can generate parodies, or ‘spoofs’ of the original text (for an analysis of different types of parodies on YouTube, cf. Willett, forthcoming).

The quoted original material can be credited in the video clip or in its paratext; in this case, the quotation is referenced. In many cases, however, there is no explicit reference as to where the original material was taken from. In this case, according to the academic genre conventions, the quote is considered as plagiarism. Quite the opposite in video-interaction, where participants use and transform previous textual materials and make a large use of implicit reference, so that viewers need to draw largely on specific shared knowledge to attribute further layers of meaning to the signs in videos:

like ‘memes’, ‘viruses’ and ‘spreadability’ with any explanatory power, it is necessary to see videos as carriers for ideas that are taken up in practice within social networks, not as discrete ‘texts’ that are ‘consumed’ by isolated individuals or unwitting masses – a ‘copy the instructions’, rather than ‘copy the product’ model of replication and variation. These ideas are propagated by being taken up and used in new works, in new ways, and therefore are transformed on each iteration; and this process takes place within and with reference to particular social networks or subcultures. (Burgess, forthcoming)

In this sense, online video production is similar to (post)modern forms of art, such as T. S. Eliot’s *The Waste Land* (1922) or U. Eco’s *Il nome della rosa* (1980), whose intertextual (Bakhtin, 1981, 1986; Kristeva, 1969/1980) reading requires the reader to identify all sources of (mis)quoted material in order to make (further) meaning out of all implicit references which are embedded in the text. By means of this implicitness, online video production is, like post-modern literary works, an elitist genre, which can be fully accessed only by those who share specific knowledge of its themes, topics, protagonists and histories, who can fully grasp the intertextual meaning of its texts.

Once this elitist referential practice is known by the watcher, coherence is sought for, rather than given. Hence, in video-interactive exchanges, whenever a video response seems totally non-sense or unrelated to another, the watcher is likely to assume that some implicit reference is there which needs interpretation on the basis of specific knowledge of some previous acts/facts, before considering the exchange as incoherent (or labelling the response as ‘spam’). In this rather specific sense, one may say that the Gricean cooperative principle is indeed at work in video-interaction: for those (only those) who know (You)Tubing conventions of implicit referencing, any interactional exchange may be assumed to be coherent and, in turn, any perceived incoherence may be attributed to a lack of shared (specific and implicit) knowledge on the part of the watcher.

#### **4.5 Further terminology: Video-interaction, (video-)thread and (You)Tuber**

Before concluding the chapter devoted to the theoretical framework of the analysis, it is necessary to explain briefly three labels which have been adopted here, namely, the terms ‘video-interaction’, ‘(video-)thread’ and ‘(You)Tuber’. These labels have already been introduced in the previous sections, yet, since they are neologisms

firstly introduced in the writer's research work (Adami, 2008a, b, 2009a, c, forth.), a further contextualization can usefully prevent any ambiguity that may rise as a consequence of their use.

The initial video, its responses (and the possible video-summary) create an identifiable 'video-thread', instantiating a new semiotic practice, referred here as 'video-interaction' to distinguish it from other interactive practices on the Website.

The video response option is still marginally used on *YouTube* (Benevenuto et al., 2008a) and what is here referred as 'video-thread' is only a (tiny) part of the communication generated by a video. As generally occurs, the most frequent practice is the least engaging and demanding one (in terms of effort, skills, technological requirements and risks for public exposure); so, watching videos is the overall most frequent interactivity on *YouTube* (a visible product of this interaction is the number of views displayed on the video page); analogously, to reply or give feedback to a video, written comments are preferred over video responses. As an example of this in the corpus, in November 2007, the initial 'Where Do YouTube' video (analysed in Chapter 5) recorded 375,122 views and 6,691 comments, against 792 video responses (and here the ratio is skewed towards responses, since the video request solicits them). Nevertheless, the very fact that the initial video solicits video responses (and that the video-summary includes only them in its resume of the whole interaction) supports the validity of considering 'video-interaction' as a practice in itself, instantiated in 'video-threads'.

Video-threads have distinctive characteristics, which differentiate them from other forms of online communication for which the term 'thread' is commonly used (like forum or email discussions). In video-threads, sub-levels are generally limited both in depth and in the number of responses, so that video responses usually receive no (or very few) responses in their turn. Again, to make an example, the 'Where do YouTube' video-thread has only three levels; 792 responses compose the first, while only 33 sub-responses and one sub-sub-response compose the other two (cf. Chapter 5, Section 1). This one-level-oriented structure is further fostered by the 'Play all responses' option, which enables the viewer to watch the responses to a given video but does not include sub-responses (cf. Chapter 4, Sections 2.4.5 and 2.4.6). Therefore the video-thread is rather built by the responses to a given video, than by an ongoing and enlarging exchange among the participants. Even more, as discussed in Adami (2009a), all responses address only the initial video without ever referring to (or intervening on) other responses' contribution. Therefore the interaction occurs primarily between each response and the video it responds to. Nevertheless, a secondary type of interaction does occur among responses, since each takes clearly into account the previous ones, trying to avoid repetitions while keeping the same pattern, as discussed in Chapter 4 (Section 3.1.1) and as evidenced in the great variety of representational resources analysed in the responses of the threads (cf. Chapters 5 and 6). Therefore, video-interaction structure resembles group interactions where a leader asks the participants to introduce themselves, with each

participant paying attention to previous contributions (so as to avoid repetitions while maintaining the same register) but never intervening on them (Adami, 2009a). In this light, the interactional structure is more similar to a ‘circle’ around the initial video rather than to a thread which unwinds from a video along the other contributions. Bearing this in mind, the label ‘video-thread’ is used here to refer to an instance of polyphonic video-interaction which can be *followed* by means of video response links established by the participants.

The term ‘(You)Tuber’ is preferred here to ‘user’ when referring to the participants in video-interaction, assuming that, in so-doing, they actively contribute to *YouTube* existence, rather than merely *using* its interface. In the videos which compose the corpus, participants never refer to themselves as ‘users’, while, especially when greeting each other, they often define themselves in relation to the activity of ‘(You)Tubing’, either with ‘tuber’ or ‘youtuber’ (with alternative spelling forms and capitalization practices, e.g., *utuber*, *ytuber*, *you-tuber*). Consistently with this variation in use, the bracketed form ‘(You)Tuber’ is adopted here. Incidentally, although the root ‘tuber’ may be perceived as a little ‘awkward’, its primary meaning related to ‘rhizome’ happily matches the metaphor of a (You)Tuber being a ‘network hub’.

## 5 CONCLUSIONS

The present chapter has reviewed traditional models and theories of communication, namely the coding-decoding model (Shannon and Weaver, 1949) and the inferential ones of Grice (1967; 1975) and Sperber and Wilson (1986). It has argued on their inadequacy for the description of video-interaction, in which the interlocutors’ mutual understanding of their intended meanings is not necessary for successful communication. The traditional notions of coherence and cohesion have also been discussed (Beaugrande and Dressler, 1981; Fairclough, 1992; Halliday and Hasan, 1976; van Dijk, 1985) with reference to video-interaction, in which apparently incoherent exchanges still build acceptable interactions.

On the basis of these considerations, the theoretical framework of the present work has been introduced, by reviewing some relevant notions of social semiotics (Hodge and Kress, 1988) and multimodal analysis (Kress and van Leeuwen, 1996, 2006), with a particular focus on the notions of affordances and interest, which will be used in the analysis of the process and of the texts of video-interaction. Social semiotics conceives signs as newly produced each time signifiers are used, the association between signifier and signified as always motivated on the basis of the sign-maker’s interest, and communication as always involving interpretation and transformation (rather than the interactants’ mutual understanding). Social semiotics’ take on communication offers a more apt perspective for the description of video-interaction, which is a ‘loose’, individualized and participatory form of communication, characterized by contemporary changes in communication (deriving from larger

social, economic and political ones).

Within this framework, the notion of an interest prompt-response relation has been introduced, which will be used as a heuristic in the analysis of the present work. At the level of processes, each semiotic act prompts a range of possibilities, which can be taken up by a responding act. At the level of texts, each complex sign sets a range of prompts. Within this range, the next text in the exchange can respond by taking up – and thus actualizing – any of the possible prompts, according to the sign-maker's interest and to the resources available to her at that given time.

The chapter has discussed the categories used in the analysis. It has defined 'video-interaction' as 'a process of exchange of (at least) two homogeneous texts (i.e., videos), one of which is a response to another' and its unit of analysis as 'composed of (at least) two videos (and the link between them): the prompting one and its video response', which constitutes the basic syntagm of video-interaction. The unit of analysis within this syntagm, i.e., between the texts of video-interaction, has also been defined, as 'constituted by any element in the prompting video and its corresponding element in the response'. Finally, some further terminology has been introduced, namely, the labels 'video-interaction', '(video)-thread' and '(You)Tuber'.

On the basis of the theoretical framework discussed so far, the next chapter introduced the methodology which has been adopted in the research.

## CHAPTER 3 METHODOLOGY

*'The map is not the territory.'*

G. Bateson, *Steps to an Ecology of Mind* (1972)

The present chapter illustrates the methodology adopted to select, transcribe and analyse the data of the study. It also discusses the sensitive issue of ethics, implied in the analysis and presentation of the material.

A first section introduces some well-known 'thorny' issues involved in analysing online data, namely those of 'representativeness' and 'significance' (1.1), of, 'stability', 'reproducibility' and 'verifiability' (1.2), and of 'storability' (1.2.1). The second section focuses on the criteria which have driven the selection of the data, both for the analysis of the interactional process (2.1) and for that of the texts (2.2); the corpus of texts is then briefly introduced quantitatively and qualitatively (2.2.2). The third section discusses the problematic of transcribing dynamic images. After reviewing some existing transcription methods (3.1), it illustrates the rationale for the *ad hoc* transcription which has been devised for the present work (3.2). The fourth section focuses on the method of analysis and, in particular, on the 'funnel' process which has been adopted in order to give account of the interactional process and its texts, so as to highlight both regularities and variations among the many data of the corpus. The fifth section is devoted to ethics. It explains the ethical standpoint adopted in the study concerning both the research process (which has been carried out through covert observation) and the presentation of the results (in terms of any potential disclosure of identity information which may derive from the description of the data).

### **1 DATA RETRIEVED ONLINE: THORNY ISSUES**

Every research needs to carefully build a rationale for the selection of the data, so as to construct a corpus which is as representative, balanced and stable as possible, for the analysis to be significant, reproducible and verifiable. By reviewing briefly these cornerstone criteria of corpus linguistics (Biber and Finegan, 1986; Facchinetti, 2007a, b; Greenbaum, 1991; Hundt et al., 2007; Thompson and Hunston, 2005), the following sections discuss how they can be hardly complied with in collecting a corpus of online data in general and of video-interactions on *YouTube* in particular.

## 1.1 Representativeness and significance

Every corpus of data collected online inevitably brings forth the question of how to select a sample which is representative of the whole phenomenon so as to be able to obtain statistically significant results. However, as is well-known,

the internet is a conglomerate of “messy data” whose size, composition and provenance constantly changes and which simply cannot be properly assessed. (Hoffmann, 2007: 69)

Therefore, the selection of a sample of online data can hardly follow a criterion of representativeness, given that one can never know (a) the amount of the total population of data existing online (i.e., the entity of the Web, but also of *YouTube* videos<sup>30</sup>), or (2) the extent of variation of this population. Consequently, no result from a sample of online data can be considered as statistically significant for the whole phenomenon. Indeed,

[w]ithout representativeness, whatever is found to be true of a corpus, is simply true of that corpus – and cannot be extended to anything else. (Leech, 2007: 135)

This constitutes a hardly surmountable limitation for a study which aims at investigating the patterns of regularity and variation of video-interaction.

## 1.2 Stability, reproducibility and verifiability

A cornerstone of any corpus-based analysis (Biber and Finegan, 1986; Facchinetti, 2007a; Thompson and Hunston, 2005) is that the corpus be stable, ‘at least for the duration of the data acquisition’ (Lüdeling et al., 2007: 9), but ‘also in the long term, so that experiments can be replicated by other researchers’ (Lüdeling et al., 2007: 9).

The ephemerality of online data, i.e., the ever-changing situation of the Web, makes collecting them particularly hard (one should virtually ‘mirror’ as much data as possible in an instant, since every second the situation is potentially different). Moreover this ephemerality does not enable the analysis to be reproduced and the results to be verified (or, at least, verifiable)<sup>31</sup> by other researchers.

Verifiability is a cornerstone of responsible research: evidence for any claim or conclusion must be subject to inspection and alternate analysis by other researchers. The web’s volatility diminishes its credibility for research. (Fletcher, 2007: 37)

---

<sup>30</sup> In this case, we need to rely on the figures declared from time to time by the *YouTube* editors.

<sup>31</sup> Indeed, even in distributed offline corpora, analyses are seldom replicated (on comparable corpora) or verified (on the same data) by other researchers; however, the possibility of doing so should at least foster a responsible research (although errors may always occur and remain undetected).



To grant verifiability, Fletcher advises that ‘Web pages on which an analysis is based must be preserved and shared’ (2007: 38) and that ‘investigators should make all webidence accessible to others for verification or reuse’ (2007: 38).

### 1.2.1 Non-storability of *YouTube* materials

In sum, for verifiability to be assured, online data need to be downloaded, Web pages need to be mirrored and the analysis needs to be carried out on the so-constructed offline database, which needs to be available to the research community. However, this solution is not viable for data on *YouTube*, since their copyright policy prohibits to download any material from the Website, as stated in the ‘Terms of Use’ page:

Content on the Website [...] may not be downloaded, copied, reproduced, distributed, transmitted, broadcast, displayed, sold, licensed, or otherwise exploited for any other purposes whatsoever.

(<http://www.youtube.com/t/terms> Retrieved 21 January 2001)

*YouTube* videos may be accessed only ‘through the normal functionality of the YouTube Service’ and for ‘streaming’, which

means a contemporaneous digital transmission of an audiovisual work via the Internet from the YouTube Service to a user's device in such a manner that the data is intended for real-time viewing and not intended to be copied, stored, permanently downloaded, or redistributed by the user. Accessing User Videos for any purpose or in any manner other than Streaming is expressly prohibited.

(<http://www.youtube.com/t/terms> Retrieved 21 January 2001)

Therefore *YouTube* videos may only be accessed online. As a consequence, if – say – a video is removed, if a (You)Tuber’s profile is hacked<sup>32</sup> or if *YouTube* servers crash, data are no longer accessible for research.

Of course, as scholars of logics and linguistic modality teach us, what is permitted and what is actually possible belong to two different realms, so that, searching the Web, it is easy to find a number of freeware flv-video downloaders, i.e., software tools which download videos produced in Flash technology (‘.flv’ is the extension of Flash files), such as *YouTube*’s. Indeed, as widely discussed and evidenced in the next chapter, the affordances of any medium are exploited by different practices

---

<sup>32</sup> This has happened to a popular (You)Tuber, *MadV*, whose video ‘One world’ has long been the Most Respected of All Times (and was intended to be included as data in the present work). When monitored on 7 October 2007, the video had reached 2,245 responses; but, at some point, its profile was hacked, as declared on its revised video description (retrieved on 31 March 2008):

This was at one time the most responded video EVER.

Sadly, someone managed to get into my account and delete all the videos, and since then all of the responses have gone.

Feel free to re-post our response if you want to feel part of the project.

On 31 March 2008, it had regained only about 50 responses. Eventually *MadV* removed the video from the channel (datum retrieved on 21 January 2009).

according to diversified interests. Hence, if viewers want to have a specific *YouTube* video available offline – in order to, e.g., watch it anytime or use some of its resources to produce their own video – they need to download it; thus some have created and shared downloading tools for these purposes. However, academic research cannot (declare to) do what is (possible yet) prohibited, even if it were the only viable solution for collecting a reliable corpus of data.

---

In consideration of the above, the data selected for the present research cannot be said to be representative of the whole phenomenon of video-interaction, nor can their results be considered as statistically significant. Furthermore no offline database can be made available to either reproduce the analysis or verify it. Finally, the data analysed here are the image of a ‘moment’, i.e., of when the data were retrieved, which is by no means the same as what is online at – say – the moment of writing, nor during the months when the transcription and analysis were carried out, let alone when this thesis will be read.

Without denying the above limitations, the next section discusses the criteria which have driven the selection of the data for the present work, which, I believe, mitigate the said limitations and enable the analysis to be significant and generalizable – to a certain extent, at least – to the whole phenomenon of video-interaction.

## 2 DATA SELECTION

As discussed above, the issues concerning both representativeness and significance of corpora selected from the Web are well known to linguists (Hundt et al., 2007). Of course, the above discussion refers to researches on texts, not on quantitative researches on network systems – e.g., on the number and distribution of response links<sup>33</sup> – which work on a completely different rationale and for which a random crawling of a huge quantity of data is a viable solution for granting representativeness.

In turn, being largely manually done, textual analysis needs to work on a – relatively – limited quantity of data, for which randomness cannot grant representativeness. Furthermore, here, dealing with videos rather than with ready-processable written texts, the issue related to the size of the corpus is even more acute. Indeed, highly sophisticated software tools can now almost automatically store, clear up, tag and parse written data captured from the Web (for a review of a number of recent software tools, cf. Hundt et al., 2007). Hence, for linguistic analysis, in a reasonably short time it is possible to collect a corpus as huge as it was unconceivable only few years ago. In contrast, video transcription is an extremely time-consuming task,

---

<sup>33</sup> The only quantitative study attempted so far on video-interaction is Benevenuto et al. (2008a).

which, at present, needs to be done manually (see Section 3); no software tool currently existing can indeed automatically ‘grab’ the content of a video and turn it into processable material. Therefore, although the corpus considered amounts to about 2,000 videos (cf. Section 2.2.2), which, qualitatively, could give material for years of analysis, its size is relatively small when compared to the whole phenomenon (the entity of which is not even ascertainable).

Nevertheless, a certain representativeness of the sample – and the consequent significance of the results – is granted by the criterion of *popularity* that has guided the selection of the data, both for the analysis of video-interaction as process (Chapter 4) and of the texts of video-interaction (Chapters 5 and 6). The next sub-sections illustrate how this criterion of popularity has driven the selection of the data for each part of the analysis, i.e., the process and the texts of video-interaction. A last section details further the video-threads which compose the corpus of texts.

## **2.1 The criterion of popularity in video-interaction as process**

The analysis of video-interaction as process has been carried out on data selected according to a criterion of popularity. Indeed, after presenting the introduction of the video response option and its affordances, i.e., what it enables and prevents (You)Tubers to do (Chapter 4, Section 2), the analysis of the process of video-interaction focuses on the Most Responded Videos of All Times (Chapter 4, Section 3), i.e., on the videos which start the largest video-threads currently existing on *YouTube* (which is the leading Website for video sharing and the only one currently enabling video responses). In sum, the threads which start from the most ‘popular’ videos (i.e., the ones which are mostly responded) are the largest instances of video-interaction. the focus on analyses. Apart from the value of ‘popularity’ in highlighting general trends, even if it may not be indicative of all video-interactional exchanges which take place on *YouTube*, this selection complies also with one of the parameters<sup>34</sup> highlighted in Claridge (2007: 90) for the collection of corpora for the study of interaction. Indeed, the larger the interactional thread, the higher the possibility of ‘capturing’ both regularities and variations in the exchanges.

### **2.1.1 The monitoring period of the process**

All videos appearing on the first two-pages of the Most Responded top chart (40 videos) have been monitored on five different days (19 August and 7 October 2007; 31 March, 31 May and 30 September 2008) over a 14-month period. Hence, by following the criterion of popularity, this selection considers the whole population of the first 40 most responded videos of all time for over a year.

---

<sup>34</sup> The other parameter is the international reach, which is also fulfilled for the data of the present work, since they have been collected from the international version of *YouTube* (national versions of the Website have started to be launched soon after the data collection).

As discussed in the analysis in Chapter 4 (Section 3), over the 5 monitoring days, 94 different videos have appeared among these 40 most responded videos of all time. This wide monitoring has enabled the analysis to establish a typology of the most responded videos, namely:

- a. Video requests;
- b. Prompting videos;
- c. Anomalous instances of top charted videos;
- d. Topic-related flooded videos.

For each of these four categories, one initial video has been randomly selected and the thread which is built by its responses (and the responses to these latter ones up to any existing level) has been considered as data for the analysis of the texts of video-interaction, as detailed in the next section.

## **2.2 The criterion of popularity in video-interaction as texts**

As illustrated above, the criterion of popularity which has driven the selection of the data for the analysis of the process of video-interaction has led to a classification of the types of most responded videos; this, in turn, has driven the selection of the data for the analysis of the texts of video-interaction. Specifically,

- a. as representative of the video requests, the ‘Where Do YouTube?’ initial video – and its thread – has been selected;
- b. as representative of the prompting videos, the ‘Best Video EVER!’ – and its thread – has been selected;
- c. the anomalous instances are illustrated by the case of the initial video ‘me getting used to my webcam’ (analysed in Chapter 4, Section 3.1.3);
- d. the topic-related flooded videos are instantiated in the analysis of the ‘Èric and the Army of the Phoenix (1/5)’ video-thread (analysed in Chapter 4, Section 3.1.4).

The initial video of the thread selected as representative of the video requests (type a. above), titled ‘Where Do YouTube?’, was charted as the 11<sup>th</sup> most responded video on 19 August 2007, 14<sup>th</sup> on 7 October 2007, 17<sup>th</sup> on 31 March 2008, 19<sup>th</sup> on 19 May and 37<sup>th</sup> on 30 September 2008. The one which starts the thread selected as representative of the prompting videos (type b.), titled ‘Best Video EVER!’, was charted as the 19<sup>th</sup> most responded video on 31 March 2008, 24<sup>th</sup> on 19 May and 39<sup>th</sup> on 30 September 2008. The ‘anomalous most responded video’ selected (type c. above), the one titled ‘me getting used to my webcam’, was charted as the 18<sup>th</sup> most responded video on 19 August 2007 and 30<sup>th</sup> on 7 October 2007; then it was removed (and its uploader account terminated) as a result of the policing practices by (You)Tubers (cf. the discussion in Chapter 4, Section 3.1.3). Finally, the selected topic-related flooded video (type d. above), the one titled ‘Èric and the Army of the

Phoenix (1/5)', has been charted as the first most responded since 31 March 2008 and still is at the moment of writing.

To the video-threads in a. and b. (initiated by a video request and a prompting video respectively) are devoted two separate chapters of textual analysis (cf. Chapters 5 and 6). In turn, due to their nature, the video-interactions in c. and d. above are not subject of a detailed textual analysis and are rather analysed within the discussion of the typology of their initial video (Chapter 4, Sections 3.1.3 and 3.1.4 respectively).

The selection of the threads in a. and b. has not been totally random. Indeed, out of all the different types of video requests (cf. Chapter 4, Section 3.1.1), the initial video of the first thread ('Where Do YouTube?') has been randomly selected among the topic-specific video requests. This selection has been motivated by the aim of exploring the textual relatedness of the responses to a quite definite and specific prompt, i.e., a topic-specific request for information (as the title of the video suggests). In turn, out of all the prompting videos, a particularly vague one has been selected, so as to see, by contrast, how relatedness is shaped in the responses to an apparently 'meaningless' video (indeed, the 4" long 'Best Video EVER!' features a face filmed while blinking twice at the camera). Hence the selection of these two particular threads has been driven by the aim of investigating regularities and differences in the patterns of relatedness when a well-defined range of prompts vs. a vaguer one is set by the initial video. As the analyses show (cf. Chapters 5 and 6), this selection has led to interesting results. Indeed, although significant exceptions are also attested, the first thread generally follows more coherent (and 'unmarked') patterns than the second one. The 'incoherent' exceptions of the topic-specific thread become the 'rule' in the second one, the one started by a vaguer video. Since, in the top chart, the incidence of (topic-specific) video requests has decreased during the monitoring period, in favour of more generic videos (cf. Chapter 4, Section 3.1.1), this apparent 'incoherence' can be said to be a distinctive feature of how relatedness is constructed in video-interactional exchanges.

There is more; as a side-reference, another thread has been considered, i.e., the one started by the video 'Why Do You Tube?' (20<sup>th</sup> Most Responded Video on 19 August 2007; 28<sup>th</sup> on 7 October 2007; later removed by its uploader). The thread has been selected to compare the multimodal deployment evidenced in the 'Where Do YouTube?' one. The selection has been motivated by (a) the similarity in topic of the two video requests (i.e., one asks for the reasons for (You)Tubing, the other for its location) and, simultaneously, by (b) their difference in the semiotic mode selected to express the topic request (the 'Why Do You Tube?' one employs speech, while the 'Where Do YouTube?' one uses writing). The comparison shows that attuning in mode is a device which builds coherence in the thread even when topic-relatedness is absent (cf. Chapter 5, Section 3.2) and hence that 'form' is often more regarded than 'content' for the success of video-interactional exchanges.

### **2.2.1 The time-span of the collection of the texts**

The data of the ‘Where Do YouTube?’ thread and of the ‘Why Do You Tube?’ one were retrieved on 19 August 2007; the data of the ‘Best Video EVER!’ were retrieved on 4 April 2008. The data of the threads started by the videos ‘me getting used with my webcam’ and ‘Èric and the Army of the Phoenix (1/5)’ were retrieved on 7 October 2007 and on 1 June 2008 respectively. The threads in the corpus were collected on different days because their initial videos appeared in the top chart in different moments during the monitoring period. This adding-up of data to the corpus during the monitoring period is mainly the result of the circularity process of the research (i.e., it is realized that further data are needed only during the analysis of the ones already collected). Although this varied time of collection may not grant the maximum synchronicity among the data of the corpus, this (short) time span provides two advantages to the research, namely:

1. the data can account for the ongoing development and change of the semiotic practices with which the video response functionality is used by (You)Tubers;
2. consequently, the present work can examine a broader range of types of interactions than would have been possible by collecting the data on a single moment in time.

Indeed, while the video request (type a. above) and the anomalous most responded video (type c. above) of the corpus have appeared on the top chart since the very first monitoring day, the prompting video (type b. above) has been charted only starting from the second monitoring day, and, even more significantly, the topic-related flooded threads (type d. above) have started to be charted in a considerable number only since the later days of the monitoring. Even more, as discussed in Chapter 4 (Section 3.1.1), in the earlier stages of the monitoring, (topic-specific) video requests played a major role in the top chart, so that, following the popularity criterion, these would have had the only ones to be considered for inclusion in the corpus. In turn, along the 14-months of monitoring, the other types of most responded videos have considerably increased in number, so that their incidence in the top chart distribution has augmented. Therefore, in the first place, it would not have been possible to detect this fourfold typology of the most responded videos if data were collected on a single moment in time; in the second place, the changing incidence of these four types of interactive practices could not have been thoroughly investigated otherwise.

### **2.2.2 The corpus of texts**

All considered, the corpus of the threads used for the textual analysis discussed in Chapters 5 and 6 is composed of a total of 1,949 videos, as detailed in Fig. 1. Of the other two, the ‘me getting used to my webcam’ thread recorded respectively 425 and 341 responses on the first two monitoring days, while the ‘Èric and the Army of

the Phoenix (1/5)’ was composed of 1,280, 5,665, 6,616 and 8,340 responses on the last four monitoring days. As detailed in Chapter 4 (Sections 3.1.3 and 3.1.4, respectively) only an exemplary part of these responses were considered for the analysis; hence the reason why these threads are not detailed in Fig. 1.

<b>‘Where Do YouTube?’ video-thread</b>	<b>Number of videos</b>		<b>837</b>
	Initial video	1	
	Responses	792	
	Sub-responses	33	
	Sub-sub-responses	1	
	Video-summary	1	
	Responses	8	
	Sub-response	1	
<b>‘Best video EVER!’ video-thread</b>	<b>Number of videos</b>		<b>755</b>
	Initial video	1	
	Responses	613	
	Sub-responses	130	
	Sub-sub-responses	11	
<b>‘Why Do You Tube?’ video-thread (side-reference)</b>	<b>Number of videos</b>		<b>357</b>
	Initial video	1	
	Responses	356	
<b>Total number of videos</b>			<b>1949</b>

Fig. 1. The composition of the corpus for the textual analysis.

Videos constitute the main data of the analysis, however all information on their pages (i.e., the overall layout, the video description and title, the comments etc.) and other related links (such as the channel page of their uploaders) have also been collected to be used as punctual side-reference for the analysis.

---

To recapitulate, the criterion of popularity has driven the selection of the data for both the process and the texts of video-interaction. For the process, popularity has meant considering the first 40 most responded videos over a 14-month period. A typification of these has led to the selection of one video-thread for each typology of initial videos. Selected among the most responded videos, these threads are among the largest instances of video-interaction currently existing, so that they give enough material in interaction to investigate the possible widest range of both regularities

and variations in the exchanges. Eventually, a further thread, always started by one of the most responded videos, has been used as a side reference for punctual comparison with one of the main corpus.

### 3 THE TRANSCRIPTION

When dealing with multimodal dynamic materials (i.e., videos), the transcription of data is a key issue (cf. Baldry and Thibault, 2006b). Many features can be singled out – e.g., language, body and facial expressions, colour, soundtrack and noises, background, camera position, pictures and drawings – and various levels of delicacy can be adopted for each feature. For example, as represented in Fig. 2 and further discussed in Chapter 4 (Section 1.1), language can be recorded in relation to its mode – whether it is spoken or written – or a deeper transcription can specify further features. Each of these features can contribute to a certain extent to the meaning made in videos, e.g., a duct tape sealing the mouth of a (You)Tuber with the handwritten location does not only answer the question ‘Where Do YouTube?’; indeed, the type of writing support is also the signifier of a meta-statement which refers to the initial video’s stated no-talking rule (cf. Chapter 5, Section 3.1.4). The same and possibly more levels of delicacy can be posited for all other modes employed in videos, which need also to be accounted for the meaning which is made through their combination and intertwining (cf. Chapter 4, Section 1.1) and, even more, for their multimodal deployment through time (cf. Baldry and Thibault, 2006b and their devised multimodal transcription of dynamic images). Hence, in transcribing data, the researcher is faced with many discretional choices to make.

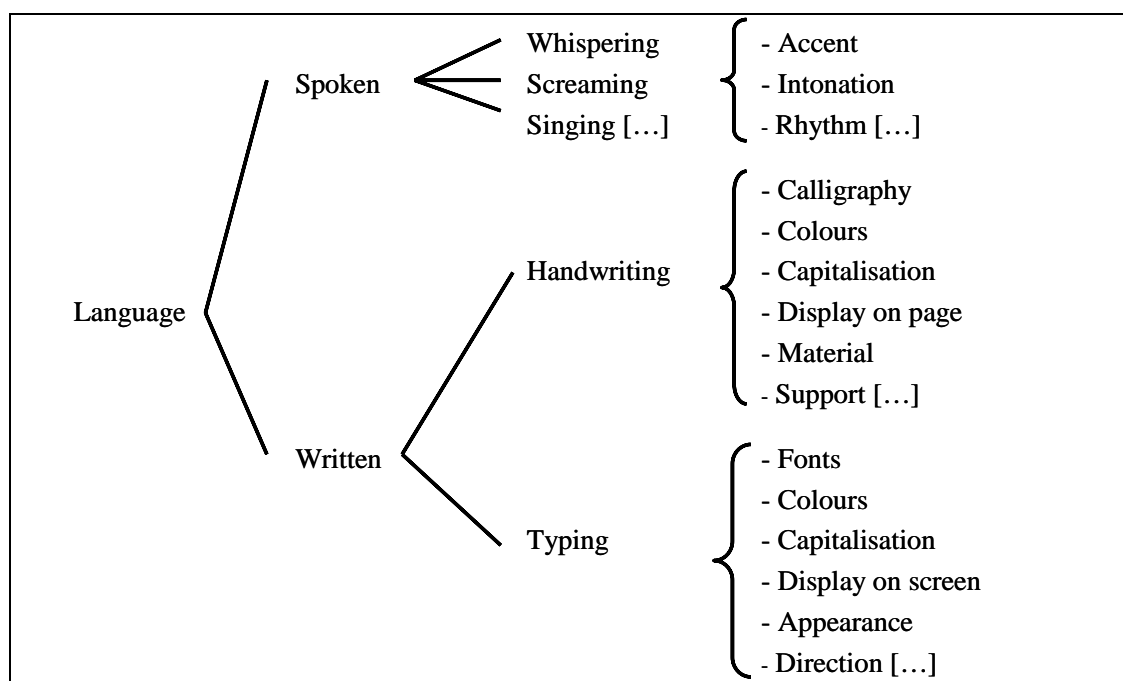


Fig. 2. Levels of delicacy in the transcription of verbal language.



Furthermore, the corpus here consists of a series of materials in *interaction*, so the transcription should not only account for the inner structure of a video (*intra-video* patterns), but also, and more importantly, for the features that may constitute regular patterns in the whole thread (*inter-video* patterns). Apparently, no transcription system has been devised yet for this type of multimodal interaction.

The next sections review some existing transcription practices for multimodal data and, specifically, for dynamic images; by evidencing the disadvantages implied in the adoption of any of these for the aims of the present work, the discussion introduces the rationale which has led to devise an *ad hoc* transcription for the data, illustrated in the last section.

### 3.1 Transcription practices for multimodal materials

Various transcription practices have been developed by the disciplines which focus on separate semiotic modes. Among these, the most thoroughly described are undoubtedly the transcribing conventions for speech used in conversation analysis, devised by Jefferson (Sacks et al., 1974). However, Jefferson's conventions for speech and turn-taking in conversation – analogously to those used for the transcription of other modes (cf., for example, the notation system used in music) – are too fine-grained for the purposes of the present work. Indeed, even a few seconds of a video can deploy all modes together and adopting a transcription system specific for each mode would not only require a huge amount of time but would also result in a hardly readable transcription, in terms of the meanings made by the combination of multimodal resources in a video, let alone the meanings made by the interaction among (several hundreds of) videos in a thread.

In recent times, various transcribing methods have been devised for multimodal analysis, i.e., for the analysis of the meaning which is made through the deployment of more than one mode. In general terms, also when they consider their material as a whole and not with a specific focus on the investigation of a single mode, different theoretical approaches transcribe (and analyse) their data in different ways. The work edited by Van Leeuwen and Jewitt (2001) collects examples on the transcription of different approaches to visual analysis, from content analysis to visual anthropology, from cultural studies to iconography, from social semiotics to ethnography. Each approach adopts a different transcription of the data on the basis of both their theoretical stance and on the focus of their analysis. Even more recently, Flewitt et al. (2009) discuss thoroughly the issue of transcribing multimodal data and review various multimodal transcription practices, such as those adopted by Hampel and Hauck (2006), Lancaster and Roberts (2007), Baldry and Thibault (2006a), Norris (2006), Goodwin (2007), and Jewitt and Kress (2003). These studies range from 'applied linguistics, visual ethnography, symbolic representation and computer mediated communication' (Flewitt et al., 2009), up to social semiotic multimodal analysis. As the authors point out, 'it is the research interests that determine the

choice of transcription' (2009)

Indeed, even when it is not stated explicitly in the literature, whenever a researcher sets herself to transcribe some material, she realizes immediately that the type of transcription needs to fit both the theoretical perspective and the purpose of the research. Furthermore, as Lemke argues,

*transcription* is not just a boring task to be left to someone else (if you can afford to pay them), but the place where theory meets data head-on and multimedia materials are re-framed for analysis in the way that *you* decide. (Lemke, 2006: xi)

In the present research I was interested in giving an account of the interaction of the meaning created by videos replying one to another, i.e., of the prompt-response patterns established by each exchange and on the regularities and variations of these patterns in the whole thread.

### 3.1.1 Dynamic images

Some scholars, such as Baldry and Thibault (2006b), and Norris (2004), have devised methods for the multimodal transcription of videos. However, these have been proved unsuitable for my investigation on video-interaction, essentially because they are – again – too fine-grained for my purposes. Baldry and Thibault (2006b) have built a software tool – *Multimodal Corpus Analysis* or *MCA* – and made it publicly available<sup>35</sup> (cf., also Baldry, 2004). Their transcription method is based on 'shots' as the primary unit of analysis of a video; shots are distinguishable by what they call 'transitivity frame' (Baldry and Thibault, 2006b: 122). For each shot, the time is recorded and the characteristics of each mode deployed in the shots are transcribed (cf. also Thibault, 2000). This transcription method is devised in view of conducting a corpus-based analysis on videos, so that for each feature a comparison among videos can be made. Intertextually speaking, a comparison of a corpus of videos puts videos in interaction among them. This could then be used for the videos of video-interaction (although the latter are *purposefully* made for the interaction; cf. the difference between 'interaction' and 'intertextuality' discussed in Chapter 4, Section 2.2). Unfortunately, as far as I could ascertain, the software tool, which enables the transcription at a high level of delicacy for each feature, does not allow for great customization. The unit of analysis, the categories and the features to be transcribed are pre-given by the software template and there is no way, for example, to record the supports and materials used in writing, nor – say – the direction and appearance of writing material. As Ochs observes,

Transcription is theory: the mode of data presentation not only reflects subjectively established research aims, but also inevitably directs research findings. (Ochs, 1979: 46)

---

<sup>35</sup> Accessible online, upon registration, at <http://mca.unipv.it/> (retrieved 21 January 2009).

Therefore, by using pre-given categories and coding systems the data transcribed are inevitably represented according to the theoretical assumptions which are on the basis of those categories and systems. In the present research the unit of analysis is not constituted by the shot, but, at the level of the process, by the interactional exchange (cf. Chapter 2, Section 4.1.1), and, at the level of texts, by any prompted element in the initial video which is corresponded in the video response (cf. Chapter 2, Section 4.1.2). Moreover, as they also admit, Baldry and Thibault's transcription is extremely time consuming, so that it is not apt for dealing with large numbers of videos (and my corpus is made of thousands of them).

Another transcription method is devised by Norris (2004) for the analysis of real-time multimodal interaction recorded in videos. Norris takes the 'action' as the unit of analysis and builds her model on multiple transcriptions, one for each semiotic mode. In her study, videos are not the textual materials – the data – but rather the recording of the data that the researcher uses in order to keep track of the interaction she wants to transcribe and investigate. Differently, in video-interaction, videos are the data in themselves; thus, they make meaning through resources which Norris does not contemplate (which go beyond the 'action'), e.g., filmic cuts, colour and editing effects and camera position/angle movements, not to mention all the disembodied resources used in videos, such as software-produced artefacts, which are normally not deployable in face-to-face interaction. Thus, although Norris' categories are more customizable than Baldry & Thibault's, they are suitable for describing what happens *in* a video, i.e., the meaning that is made by the interactants in real-time (captured by the researcher through a video), not for the meaning that is made *by* a video interacting with another one in asynchronous communication. Furthermore, also here the level of delicacy of the transcription is too deep; indeed, for each video, multiple separate transcriptions are made, one for each semiotic mode for each unit of analysis, i.e., each action. Therefore, also this transcribing practice makes a high number of data hardly manageable.

It must be noted that, in case of a high number of data, the disadvantage of a (too) fine-grained transcription is not due merely to the fact that it is time-consuming. A highly detailed transcription becomes also a less-manageable tool for the analysis of a large corpus of data. Indeed, when the object of the analysis is, as in Thibault (2000), a 60'' advertisement, a fine-grained transcription – as time-consuming as it may be – is desirable, since the more the transcribed elements (and relations among them), the more thorough the analysis of the meaning of the ad. Conversely, when the analysis focuses on thousands of videos (or several hundreds of them, considering each thread of the corpus separately), a too fine-grained transcription makes the transcribed data rather cumbersome to manage. When videos are numerous, a transcription needs to carefully balance detail and the 'wholeness' of the thread, so as to enable the analyst navigate easily through it, easily retrieving each feature, easily identifying each video and possibly, avoiding the need of having to re-watch the original material every time a feature is analysed. As detailed in Chapter 2 (Sections 4.1.1 and 4.1.2), the unit of analysis of video-interaction is the video-

interactional exchange composed of two videos (the initial one and its response); it is a higher level unit than both Baldry and Thibault's (2006b) – i.e., the single shot of a video – and Norris' (2004) – i.e., the single action in a video.

### **3.2 The *ad hoc* transcription devised for the data**

On the basis of the above considerations, the present section describes the rationale which has grounded the *had hoc* transcription devised for the data of video-interaction.

#### **3.2.1 An interest-driven selective transduction**

From the above discussion it becomes apparent that every transcription is a selective transduction of a video. The type of selection (the elements of the video that are transcribed, i.e., transduced) depends (1) on the type of material (i.e., hundreds of videos in a thread, in this case) and (2) on the research question (i.e., how they relate to each other and to the initial video). In other words, like all forms of representation, also transcription is an interest-driven transformation of existing resources, which responds to a prompt (the type of material) according to an aim (the research question). What I needed was a transcription that could enable me to highlight the patterns of regularity and variation that occurred in the interaction among videos, rather (or more) than for the analysis of each single element or mode within a single video. And no transcription has so far been devised for videos that interact one with another, essentially because this form of communication is newly-born.

#### **3.2.2 A cyclic process driven by recurrence, saliency and relevance**

The method employed in the transcription of the present data is empirical and subjective. At the very first stage of the observation, I have simply started to watch the videos and tried to single out what I was noticing both as salient and as recurrent. Obviously, the more the watched videos, the more the salient and recurrent features to be added, so that, each newly added features has required watching all the already-transcribed videos several times until the last video in the data was transcribed<sup>36</sup>.

Transcribing data is inevitably a cyclic process which involves also a first analysis of the data that are being transcribed<sup>37</sup>. On the basis of this first analysis, new features can be highlighted as relevant and can thus be added in the transcription. In all cases, what is utmost important, for any transcription to be reliable, is to be maximally

---

<sup>36</sup> For example, in the 'Where Do YouTube' thread, at the 427<sup>th</sup> video, I chose to transcribe whether any opening or closing gestures were enacted in front of the camera, since I realized it was a quite common, though diversified practice; the analysis of this datum can be found in Adami (2009a).

<sup>37</sup> Cf. also Goodwin: 'different stages of analysis and presentation will require multiple transcriptions. There is a recursive interplay between analysis and methods of description' (2001: 161).

consistent among the transcribed data (which is not a straightforward task to be achieved, especially when the data are many and their transcription requires several months of work). Furthermore, being selective, any transcription never substitutes the original data and whatever feature is analysed on the transcription must be continuously checked against them.

Eventually, the features noted down in the transcription of the corpus – as said, subjective and empirical – are motivated by:

- a. *relevance* (i.e., what I was searching for);
- b. *saliency* (i.e., what has stricken me while watching);
- c. *recurrence* (i.e., what I have noticed as common in videos).

Needless to say, other research purposes would require other types of transcription. In sum, I have chosen to adopt an *ad hoc* transcription for my data (‘*ad hoc*’ is used here in its Latin sense, as ‘specific for the task’). However, as subjective as it may be, the two criteria of recurrence and saliency enable the transcription to trace respectively both the regularities and the variations that occur among videos.

### **3.2.3 The transcribed (recurrent, salient and relevant) features**

As said, subjective and empirical, motivated both by what I was searching for and by the features perceived as salient and recurrent while watching, the transcription accounts for the following features:

- Video data-logs;
- Paratextual info
- The ‘content’ (i.e., the ideational elements);
- The mode(s) in which the content is produced;
- Presence/absence of the representation of the (You)Tuber;
- The background;
- Opening/closing signs;
- Sounds (soundtrack/environment noise);
- Colour;
- (my personal) notes.

Firstly, I have noted down the identifying details of every video: its URL on the Website; its positioning in the thread (the response number); the time of duration; the date when the video was posted; the username of its uploader. Then follow paratextual information of the video: its title; what is displayed on the video thumbnail chosen by the uploader to (re)present the video file on the responses page; the video description (found in the ‘About this video’ section on its page); the number of comments (transcribing the significant ones) and of video responses that were posted to the video. The transcription of the video content accounts for the

ideational meaning that is displayed in each shot (the represented ‘who’ and ‘what’) and the semiotic mode through which it is represented. I have also noted whether the video displays the protagonist’s face (or avatar), the background shown in the video, possible opening and/or closing signs of greetings (with their related semiotic mode of representation), the type of sounds (whether there is music or environmental noises in the background), if any, and the colour effects, broadly distinguishing whether the video is shot in black and white or in colours and, in this latter case, whether a predominant colour is employed. Finally, a ‘notes’ column records any salient feature that does not fit in any of the other categories.

### **3.2.4 A ‘monomodal’ transcription of multimodal texts**

Although Flewitt et al. (2009) rightly stress that a multimodal analysis should preferably adopt a multimodal transcription, the one I have devised is mainly done in type-writing on a database worksheet arranged in rows and columns, with hyperlinks directing to the original video page on *YouTube*. This essentially monomodal transcription has been chiefly motivated by the resources available to me and by the need of reducing the time required to transcribe the data (which would have considerably increased if I were to include in the worksheet also images of each video shot). Still, it has also been motivated by the need of easily handling the huge amount of data. Indeed, unlike images and colours (e.g., using different colours for – say – signalling different modes deployed in videos), digitalized typewriting offers the undeniable advantage of enabling the automatic processing of the transcribed data, so that any feature can be searched, ordered and counted, and can be turned into data represented in graphs and diagrams. Given the large number of videos, the possibility of handling the data also quantitatively is definitely important in order to account of both regularities and variations in the threads.

Eventually, these transcribed worksheets have become the basis of the textual analysis of each video-thread and further columns have been inserted according to the diversified focuses at each stage of the analysis (e.g., the list of the declared locations and the textual organization of the representations for the ‘Where Do YouTube?’ thread; the type of relatedness for the ‘Best Video EVER!’ thread).

## **4 THE METHOD OF ANALYSIS**

As evidenced in the discussion so far (cf. 2 and 3) the analytic process begins and includes both the selection of the data and their transcription. Indeed, on the one hand, the selection of the texts of the corpus has been the result of the analysis of the process of video-interaction and, specifically, of the determination of a taxonomy of the different types of Most Responded Videos on the Website. On the other hand, the transcription of each feature has coincided with the first analytical singling out of the salient and recurrent elements in the videos.

#### **4.1 The funnel process**

Both the selection and analysis of the data follow a ‘funnel’ process, which zooms in from a general phenomenon up to a more specific one, by examining both exemplary instances and relevant exceptions at every level. So, the analysis of the process of video-interaction starts from the affordances of the video response option and then focuses on the Most Responded Videos of all time. Thanks to a categorization of these into four main typologies, the analysis then zooms in to the video-interactive practices which are characteristic of each of them. Then, the analysis zooms-in again and examines textually an exemplary thread for each typology.

Analogously, the so-selected texts of video-interaction (the exemplary threads) are considered in their general composition according to the function that each video plays in the thread (i.e., initial video, responses, sub-responses, possible video-summary and so on); within this functional grouping, the initial video (and the possible video summary) is described in its form and analysed in its possible meanings, while the responses are analysed by categorizing them according to the type of relatedness which they establish with the initial video. Each relatedness grouping is described quantitatively and qualitatively, in terms of its general textual features; then, within each of these groupings, exemplary responses are described and analysed as instances of the whole category; exceptions are also singled out so as to give account of both the regularities and the variations of the thread.

#### **4.2 The pilot study**

The ‘Where Do YouTube?’ video-thread has constituted the data for the pilot study of the present work (Adami, 2008a, b, 2009a, c, forth.). It has firstly focused on the multimodal deployment of the thread and has compared it with the one of the side-reference corpus, started by the video ‘Why Do You Tube?’. The pilot study has identified the function of attuning played by the topic-mode – i.e., (hand) writing for the first thread, speech for the side-reference one – and has singled out some significant exceptions to traditional coherence patterns.

The analysis conducted on this pilot study has enabled the research to refine (a) the theoretical framework (i.e., the inaptness of traditional communicative models for the description of video-interaction), (b) the research question (i.e., the regularities and variations in the patterns of relatedness in the thread) and (c) the analytical tools of the work (i.e., the interest-driven prompt-response relation; newly shaped multimodal cohesive ties and attuning devices), together with (d) the first observations (e.g., patterns of relatedness which only partly conform to traditional notions of coherence and cohesion; successful interactions in the thread which are nonetheless apparently incoherent; a relevant role played by form – rather than content – in establishing relatedness in the thread), which needed to be verified on the other thread, the one started by the ‘Best Video EVER!’.

### 4.3 Quantitative and qualitative interpretation focused on signifiers

The analysis is both quantitatively and qualitatively conducted; quantity counts of recurrent elements trace some first regularities, while qualitative investigations give account of the variations and enable the categorization of the data into different typologies. Needless to say, while numbers give evidence of the import of a given phenomenon, often exceptions are indicative of highly significant trends, which may represent the instantiation of newly born practices.

As for the analysis of the signifiers of the videos, their interpretation has drawn upon multiple sources. Firstly, the results found in the literature of works on multimodal analysis<sup>38</sup> have been used, above all Kress and van Leeuwen's (1996, 2006) 'grammar' for the (Western) conventions used to make meaning in multimodal texts. Furthermore, in recent times *YouTube* videos are being subject to extensive analysis, all of which has been used for the interpretation of the data of the present work (cf. the works of Bardzell, 2007; Bruns et al., 2007; Burgess, forthcoming; Burgess and Green, 2008; Cha et al., 2007; Clemons et al., 2007; Eastment, 2007; Halvey and Keane, 2007; Lange, 2007a, b, 2008; Lee, 2008; Melican and Faulkner, 2007; Regan and Revels, 2007; Shida and Gater, 2007; Turkheimer, 2007; Weiss, 2007; Willett, forthcoming).

Beside the literature, I have drawn also on my own experience as a (Western) viewer of both mass-media videos and films and of *YouTube* videos. Therefore, my interpretation is at least partly based on my experience of (Western) filmic and online-video conventions. When deemed useful, this interpretation has been verified with any possible feedback given by the written comments to a video. Yet, not only any interpretation is necessarily subjective, but, when found in a comment, it also constitutes a further representation, which, as a consequence, requires a semiotic analysis in itself and a further (forcefully subjective) interpretation.

In a social semiotic perspective, any representation is subject to the meaning-maker's interested interpretation. This involves also the researcher, who can support her interpretation with thoroughly discussed facts but who cannot step outside<sup>39</sup> the vicious/virtual circle of the interest-driven transformation of resources which is inevitably involved in the process of meaning-making. Paradoxically, it may be said that the virtual circle of (subjectively-driven) representation-interpretation-

---

<sup>38</sup> Dynamic images have been investigated by Baldry and Thibault (2006b) and Machin and Jaworski (2006); typing has been investigated in Machin (2007) and Mavers (2007) respectively; sounds by Van Leeuwen (1999); embodiment by Norris (2004), Johnson (1987), and Gray (2004); gestures by McNeill (1992), Alibali et al. (1997), Kress et al. (2001), Heath and Luff (2007), and Sutton-Spence and Woll (1999). Gaze has been investigated by Kendon (1967), Streek (1993), Lancaster (2001), Bourne and Jewitt (2003), Tomasello (2003), and Roten et al. (2000), who also investigate body position. The latter has been the focus of the analysis in Royce (1984), in Flewitt (2005), Lancaster and Roberts (2007), and Vygotsky (1978), while Norris (2004) also examines the layout of space (i.e., furniture) as a resource to make meaning, and so do Dicks et al. (2006).

<sup>39</sup> For a discussion of the researcher as part of the researched phenomenon, cf. Bateson (1972).



transformation which makes communication possible (and so fascinating!), becomes a vicious circle when the meaning-maker is a researcher who needs to give scientific grounds to her observations.

Without trying to escape this dilemma, this is the reason why no attempt has been made to contact the authors of the videos and ask them for their intentions or interpretations of their artefacts. In a semiotic perspective, any interview would result in a further representation, in a further text which would require a specific analysis. Moreover, following the theoretical standpoint of the present work, any representation given by the author in an interview would be an interest-driven response to the prompts given by the interviewer. Consequently, this would constitute a further corpus of data to which to devote a separate analysis. Indeed, further research on the subject could well focus on the regularities and variations in the meanings produced in the video *compared with* the meanings produced in the representations given by their authors when asked about their intentions; however, this would answer a totally different research question. For the purposes of the present analysis, the almost two thousand videos (and their paratexts) constitute a sufficiently large corpus of data for discussion (even too large, so that many interesting, but secondary, observations had to be left out from the present work).

Any analysis is necessarily a selected interpretation of reality; the present one is mainly focused on signifiers, on the resources used in the videos to make meaning. True, also the signifiers discussed at each point in the analysis are the fruit of a selection on the basis of ‘saliency’, ‘recurrence’ and ‘relevance’ to the purposes of the analysis (as discussed in the section devoted to the transcription, cf. 3). However, although selective, the analysis is focused on formal elements which are indeed present in the videos, rather than on the intended meaning assigned to them by their authors. This, combined with the quantitative analysis contributes to turn the subjective reading of the data into a scientific interpretation of the phenomenon.

## 5 ETHICS

A wide debate is currently ongoing for what constitutes an ethical conduct of research on online data; for a summary of the ethical issues on Internet research, cf. Jones (2004), Knobel (2002) and McIntyre (2003), while ethical guidelines can be found in Ess (2001), Haigh and Jones (2005), Sharf (1999), and Walther (2002). Ethical concerns in academic research involve both (a) the issue of how research should be conducted so as not to do any harm to the observed participants and (b) the issue of how presenting the data so as not to do any harm in exposing the participants’ identities to the public (examples can be found in Eichhorn, 2001; Reid, 1996; Smith, 2004). The following two sections discuss the ethical stance adopted in the present work in relation to each of these issues, while a last section briefly discussed copyright issues.

## 5.1 Ethical stance in conducting covert observation

As to the first issue, I have conducted my research as a covert observer, never disclosing my presence (as viewer) and identity and aims (as researcher) to the participants, nor ever intervening and participating in the interaction. The covert observation and the avoidance of any (covert or overt) participation has prevented any influence on the data observed. This has assured the most suitable environment for the investigation of the phenomenon, in compliance with my research aims, i.e., to detect the regularities and variation in the practices of video-interaction *as it is* (and not as it would have been, as the fruit of my interaction with the participants).

Following Whiteman (2007: 75-94) and her well substantiated ‘Defence of covert observation’ (2007: 83), I have chosen this method on the main assumption that the data selected for the present study are on a public domain. The public status of video-interaction on *YouTube* is undeniable; it is well promoted by its slogan (i.e., ‘broadcast yourself’) and is also evidenced by the fact that various affordances of the medium enable participants to limit the extent of publicity of their interactions (e.g., the ‘set to private’ option which is enabled for every video uploaded on *YouTube*). Also the ‘Privacy Notice’ of the Website warns its contributors on the public exposure of their uploaded videos:

Any personal information or video content that you voluntarily disclose online (on discussion boards, in messages and chat areas, within your playback or profile pages, etc.) becomes publicly available and can be collected and used by others.

(<http://www.youtube.com/t/privacy> data retrieved on 3 February 2009)

In this regards, however, Whiteman (2007: 75-94) discusses the fact that a public space may be not perceived by its participants as public to researchers:

Perceived privacy entails that some community members “do not expect to be research subjects” (Eysenbach and Till, 2001). In Goffman’s terms, this suggests that “open, unwall’d public places” may be (mis)regarded by members of Internet communities as “soundproof regions” (Goffman, 1963: 10). In a similar way, King proposes two general issues that should be considered in relation to sites in cyberspace in order to determine how results should be reported; the nature of accessibility to the site, and the perceived privacy of members (King, 1996).

With reference to King’s issues of ‘accessibility’ and ‘perceived privacy’, in video-interaction, not only the technical affordances of the medium grant the openness of the semiotic space where the exchanges take place, but also none of the texts in the corpus lead to doubt the participants’ clear perception of openness (i.e., that they may mis-perceive that place as being private). As detailed in Lange (2008), (You)Tubers are generally well aware of the publicity status of the Website and exploit the affordances of the medium so as to variously modulate the extent of publicity of their videos (e.g., by setting them to private or by avoiding to use popular keywords or to link them to other videos, so as to prevent their videos to be accessed through either text searches or video browsing). By selecting the threads

starting from the Most Responded Videos, I have assured that the initial videos of the corpus have a highly public status – they are indeed top charted on the Website – and, consequently, that their authors do not consider publicity as a danger to be avoided. As illustrated in Chapter 6 (Section 1), the initiator of the ‘Best Video EVER!’ thread is a *YouTube* celebrity who welcomes and seeks publicity. Also the initiator of the ‘Where Do YouTube?’ thread is a quite popular *YouTube* vlogger (Chapter 5, Section 1) and nothing on his channel or video production leads to infer that he dislikes publicity. As for the respondents in the thread, by posting a video response to a top charted video, participants clearly link their videos to a very popular one and, by so doing, they increase the publicity potential of their own videos. This is generally known, acknowledged and welcome on *YouTube* and is further evidenced in the threads of the corpus by the many response inviting viewers to subscribe, ‘stay tuned’ and watch the (You)Tuber’s video production and channel. This is also evidenced, by contrast, by the few videos in the thread which are set to private and thus exploit an affordance of the medium so as to limit the access to their videos only to invited friends. Needless to say, during the research, no attempt has ever been made to access the videos which were set to private.

Therefore, unlike suggested by Bakardjieva and Feenberg (2001: 234) for other cases of covert observation, during the period of observation, I have never felt in a position of ‘spying’ some unwilling and unaware participants in their privacy. This is also in consideration of the fact that, on *YouTube*, ‘lurking is a “normal” practice’, as noticed also by Whiteman (2007: 87) in her observed fan-forums. Indeed, as is well known, on *YouTube*, viewers outnumber by far those who provide content; clearly, lurking on *YouTube* may be done for diversified purposes, such as taking some leisure time off work, writing a news article, deciding where to place a commercial advertisement on the Website, or even for research purposes.

Therefore, when I have turned from a viewer of *YouTube* videos for entertainment purposes into a viewer for research purposes, it has seemed perfectly ethical to do so without notifying the authors of the videos I was watching. *YouTube* promises their contributors to ‘broadcast’ themselves and, generally speaking, when uploading videos on the Website, their uploaders want to get as many views as possible, otherwise they use various strategies to make their videos less immediately retrievable; cf., again, Lange (2008). Daily, thousands of *YouTube* videos are embedded in other Websites and are used/promoted for various purposes; they are forwarded in email messages and talked about in newspapers and TV news broadcast and shows; why would they not be ‘ethically’ allowed to be researched?

Admittedly, when approaching the huge literature dealing with ethical issues of online research, I have started to have some doubts as to whether I had maybe naively ignored a delicate and sensitive issue which would instead deserve a more thorough reflection. Nevertheless, I have cleared these doubts quite straightforward, in the firm belief that the authors or my data are and deserve to be regarded as virtual film-makers. They produce their videos as craftily as they can, according to their

interests and to the resources available to them; they upload them in a public space which, not incidentally, promises them to ‘broadcast’ themselves; even more, they link their video responses to massively responded – and thus public – videos. By so doing, they unquestionably welcome – and sometimes explicitly ask for – their videos to have a public of viewers. Furthermore, most participants present themselves as ‘directors’ on their *YouTube* channel<sup>40</sup>. The initiators (but also many of the respondents) have their videos watched by millions of people; thus they reach an audience as large as many professional film-makers can only dream of.

Therefore, the authors of my observed data are no less than (amateur)<sup>41</sup> film-makers and they should be regarded as such. No sensible researcher would ever ask a professional film director for her permission to investigate her creation<sup>42</sup>. In this view, I believe that by asking (You)Tubers for their permission, I would have endorsed a quite patronizing attitude towards them, i.e., I would have regarded them as being of an inferior status than that of film-makers, which – I believe – they are not.

## 5.2 Ethical stance in the presentation of the data

As to the ethics implied in the presentation of the data and in the possible disclosure of sensitive information on the participants’ identities, I here again follow Whiteman (2007) in considering what represented in the videos as texts and not as people, even when these texts represent the participants’ faces filmed at the camera. Here again, my data prove that (You)Tubers are well aware of the various strategies that can be enacted so as to avoid any disclosure of personal information. The discussion in Chapters 5 and 6 on the diversified use of avatars or of various means to disguise their identity in videos, and on their avoidance of providing personal information, bring strong evidence to the fact that (You)Tubers are aware of the potential danger of disclosing personal information online. As a consequence, they choose the resources represented in their videos according to the amount of sensitive information which they are interested in disclosing.

The presentation of the data never refers the texts to the offline identity of their authors, nor the discussion ever intends to (or could) say anything about it. As discussed in Chapter 2 (Section 4.2), the research focus is on the semiotic space and the representational resources which are produced and transformed in the interactions which take place in it. No interest is ever placed in who the participants are or in what their intentions are, but rather, in how characters and personae construct and present them – and their intentions – through their representations. A face expression

---

<sup>40</sup> ‘Director’ is one of the possible statuses that can be selected within the information published on the (You)Tuber’s profile.

<sup>41</sup> For my reluctance to use the term ‘amateur’ in the present work, cf. footnote 22.

<sup>42</sup> In this case, only copyright terms of use would need to be complied with, for the use of the material for publication and public distribution. For the legal issue on property rights cf. Section 5.3.

in a video is a semiotic resource, a signifier, in the same way as an emoticon or a spoken 'I am pissed'. In this terms the represented faces are treated and analysed in the present work, i.e., as signifiers created by the sign-maker and used as signs to convey meaning. Every representation of the 'self' conveys some information on the identity, or, better, a certain representation of the identity. However, these representations are always referred to the character, to the protagonist of the video, who, when coinciding with its author and uploader, is a signifier of the latter and functions as a text produced by her. If the author is dead, as Barthes argues (1977a), the texts intentionally produced and uploaded online on a public domain cannot but be considered as mere signifiers, even if these are the author's filmed face. They can by no means be conducted to the offline identity of their author in flesh and blood.

In sum, if an acknowledged limitation of the present study is the impossibility of saying anything on the participants in flesh and blood, because it only considers their online texts, the same argument should be kept valid also for the ethics of presenting and discussing the contents of these texts, which are indeed made of signifiers rather than people.

As to anonymity, the presentation of the data does not conceal the initiators' usernames, in reason of their proved public status as *YouTube* celebrities (while, in reason of their differentiated extent of popularity, the respondents' usernames are not made public, even if they often promote their names in their videos as well as those of other (You)Tubers). The titles of the videos are useful paratextual data in the analysis and thus are subject to discussion. Through a keyword search, readers could possibly retrieve the discussed videos on the basis of their cited titles. This is not considered here as constituting any potential harm to their authors; indeed, only what is willingly published online by their authors is accessible. For the benefits of the research, this avoidance of anonymity of the data grants a certain extent of verifiability of the study<sup>43</sup>, while, to the participants, it may only grant further views to their video creations, which – as discussed earlier – they never show to dislike.

### 5.3 Copyright issues

Stepping outside the realm of ethics, it is worth mentioning some legal issues of property rights. The screenshots reproduced in the present work are taken from non-copyrighted videos, while, when images of copyrighted features of the *YouTube* interface are presented (e.g., in the case of the few screenshots which reproduce the Website layout), this is done within the limits of 'fair use'.

---

<sup>43</sup> Nevertheless, as discussed earlier in Section 1, the situation on the Website is constantly changing, so that, at the time of reading the present work, the discussed videos may have been removed or may have changed their status, including their title.

## 6 CONCLUSIONS

The present chapter has illustrated the methodology adopted for the research. It has firstly focused on the selection of the data, by discussing the difficulty in collecting any data from the Web which can grant representativeness of the sample, significance of the results, and reproducibility and verifiability of the research. More specifically in video-interaction, the above mentioned criteria could not drive the selection of the data, also in reason of the legal constraints which prohibit the storage of *YouTube* videos, and due to the materiality of the data, which inevitably require manual transcription and analysis. In order to balance the inevitable shortcomings deriving from the above, the selection of the data has been driven by a criterion of popularity, both for the material used for the analysis of the process of video-interaction and for the texts. A 14-month period of monitoring of the Most Responded Videos has enabled the analysis of the process and, simultaneously, a classification of the types of videos which has driven the selection of the texts.

Secondly, the chapter has focused on the transcription of the data, by examining some existing practices for the transcription of multimodal data, and of dynamic images in particular. The inaptness of these for the purposes of the present study has been discussed before introducing the rationale for the *had hoc* transcription devised for the texts. Criteria of recurrence, saliency and relevance have driven the transcription of the corpus, motivated both by the research interest and by the type of data.

Thirdly, the method of analysis has been illustrated. In a cyclic process, the analysis has played a major role since the very first stages of the study, i.e., the data selection and the transcription. It has also led to the refinement of the theoretical framework and analytic tools used for the final analysis. It has followed a funnel process from the higher level of the process, zooming in to the texts of video-interaction, up to the distinctive patterns of regularity and variation in the interactional exchanges. In a social semiotic perspective, as the transcription is an interest-driven transformation (transduction) of the data, also the analysis is an interest-driven interpretation of them; in order to give 'scientific' grounds to this inevitably subjective interpretation, the analysis has mainly focused on the signifiers which are present in the texts and has been carried out both quantitatively and qualitatively.

Finally, the chapter has discussed the ethical stance which has driven both the covert observation of the research and the presentation of the data. In reason of the well acknowledged public domain of the semiotic space in which the interactions of the study take place and of the popularity criterion which has driven their selection, no informed consent has been asked to the authors of the observed videos. This has granted the research an optimal environment, uninfluenced by the presence of the researcher. Strong evidence has been brought to the fact that video-interactants are to be regarded no less than film-makers and, in reason of that, any request for permission of investigating their creations would have shown a patronizing attitude

which is by no means endorsed by the researcher, who is a passionate watcher and admirer of *YouTube* videos (as well as a video-maker herself). The discussion on the ethics of the present work has also involved the data presentation. Data are here considered as texts (and not as people), which are uploaded so as to interact in a public semiotic space. When the presented texts disclose any personal information concerning the identities of their uploaders, they do it by means of signifiers, through representations which are intentionally created by their authors to be publicly viewed, used and transformed. The presentation of the data deals therefore with signifiers, which can by no means be traced back to the offline identities of their authors. Eventually, no copyrighted material has been used for the images of the videos presented in the study, while the occasional inclusion of copyrighted material – limited to a few screenshots of the *YouTube* interface – complies with the terms of ‘fair use’.

Following the here-discussed methodology, the next chapter steps into the analysis of the data, by focussing on the process of video-interaction. It investigates its structural characteristics, its affordances, how these are used, exploited and transformed in the practices, and what types of videos – and related threads – have appeared on the Most Responded Videos top chart during the monitoring period of the research.





## **CHAPTER 4**

### **ANALYSIS 1/3:**

## **VIDEO-INTERACTION AS PROCESS**

*'This is kinda cool, to communicate, isn't it? It's 'unique'!'*

Video response (handwritten excerpt)

The present chapter focuses on video-interaction as process. Section 1 analyses the distinctive features of video-interaction, so as to map its place within our contemporary semiotic landscape; section 2 discusses the introduction, the affordances and the use of the video response option, while Section 3 illustrates the introduction, the affordances and the use of the 'Most Responded Videos' top chart, so as to outline the main practices in use in the largest instances of video-interaction. Through a detailed discussion based on empirical evidence, the main argument of this chapter is that structural characteristics (distinctive features), affordances and practices influence each other incessantly; in particular, the interactants' use of the affordances according to their interests generates semiotic practices which lead to changes in the structure itself. The main function of this chapter is to introduce thoroughly video-interaction as a form of communication (as a process) by investigating how it 'works' at a general level, before analysing in detail the threads of the corpus, i.e., the texts of video-interaction (Chapters 5 and 6).

### **1 STRUCTURE: DISTINCTIVE FEATURES**

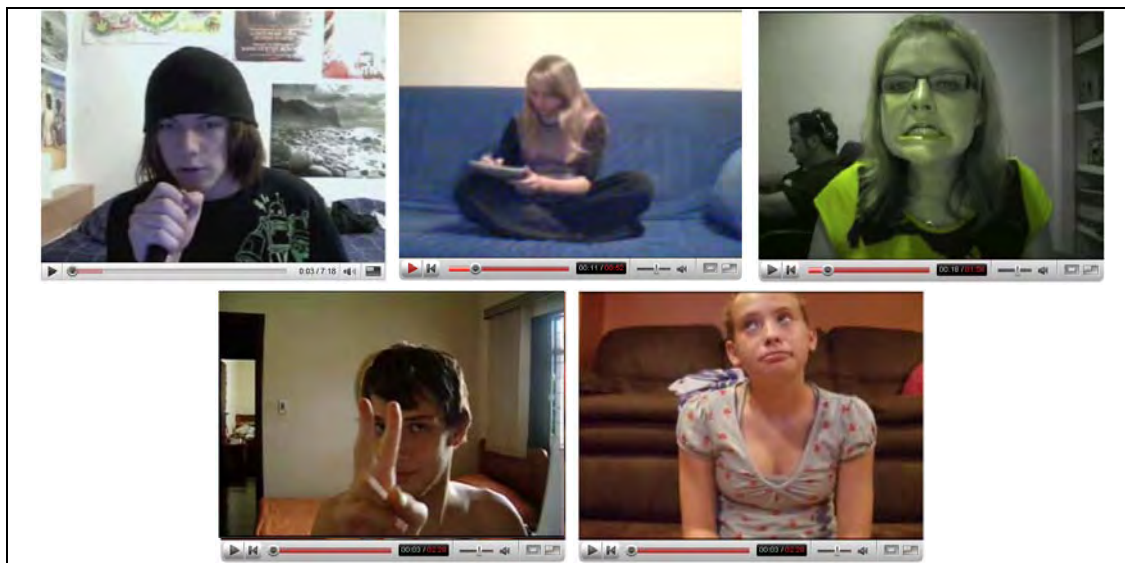
Video-interaction is a new type of communication. In order to map its place within our semiotic landscape, it is useful to identify its structural features and compare them with the ones of other forms of communication. Here it is argued that video-interaction is distinctively characterized by:

- (embodied and disembodied) multimodality;
- homogeneity and bidirectionality;
- publicity;
- asynchronicity;
- disembodiment;
- online communication;
- distance (absence of spatial co-presence);
- multiple mediation;
- corporate interface distribution.

The combination of these features makes video-interaction distinctive and determines its affordances as a new type of communication. The next sub-sections discuss each feature separately, before considering them all together and analysing their distribution in other forms of communication, so as to draw the place of video-interaction within our semiotic landscape (1.10).

### 1.1 (Embodied and disembodied) multimodality

Videos are texts which are highly multimodal, in that they afford a wide range of semiotic modes to be deployed, both in simultaneity (arranged in space) and in succession (arranged in time). Videos can make meaning through both embodied and disembodied resources (Norris, 2004).



**Fig. 3 Embodied resources in videos (from left to right, top to bottom): speech, body posture, facial expression, gesture and gaze.**

Embodied modes have the human body as their essential medium of production; so, as exemplified in Fig. 3, signs can be produced in a video by means of:

- spoken language;
- body posture;
- gestures;
- facial expressions;
- gaze.

Disembodied modes are the ones which need other materials and supports to be produced and result in semiotic artefacts (i.e., texts-as-objects), separated from their producer. As exemplified in Fig. 4, disembodied resources in videos include:

- hand- and type-writing;
- soundtrack and noises;
- drawings;
- photos;
- animation;
- the layout of – say – furniture,
- clothing.

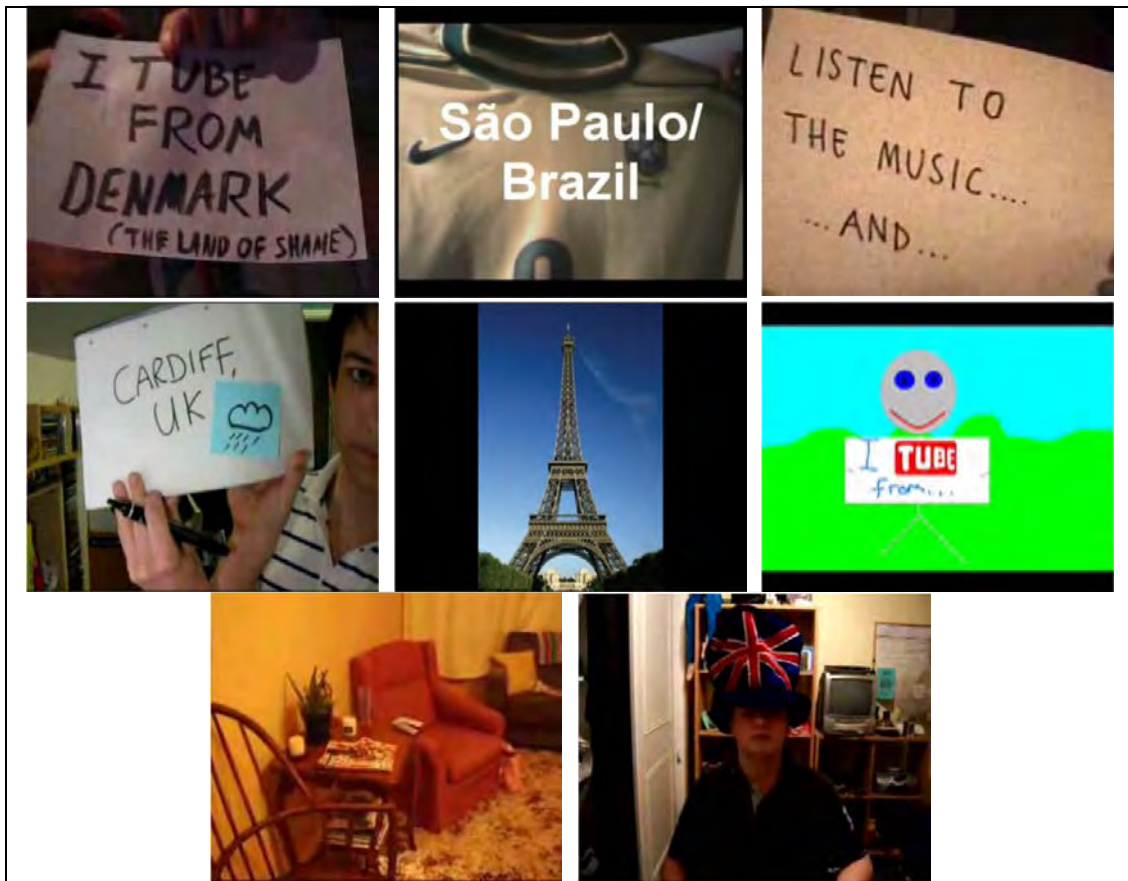


Fig. 4 Disembodied resources in videos (from left to right, top to bottom): handwriting, typewriting, soundtrack, drawing, photo, animation, furniture layout and clothes.

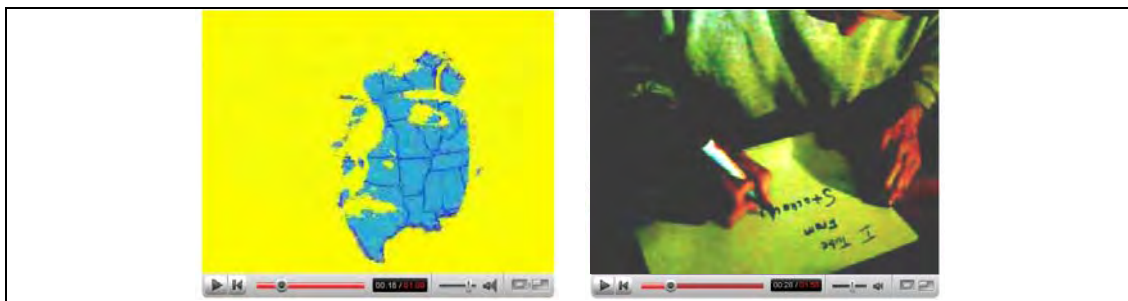


Fig. 5 Meaning made in videos by means of colour effects (left) and camera position (right).

Besides, as exemplified in Fig. 5, videos can make meaning through the resources which are specific of dynamic images, such as colour effects, the camera position and filming cuts.

Sign-makers can choose among this wide range of resources, according to their interests and their availability (i.e., hardware and software tools for video-making and the skills in mastering them). Therefore, every sign which is used in a video is the fruit of a selection from a ‘paradigmatic’ axis (Saussure, 1931)<sup>44</sup> of available resources, so that it is not (only) significant *per se*, but it is (also) significant in that it has been preferred (considered as the most apt) among a range of options.

Obviously, each of these semiotic systems (i.e., modes) has sub-features which contribute to shape meaning. For example, spoken language can be whispered, screamed or sung, with various accents, tones, rhythms and intonations. In turn, writing can make different meanings through capitalization, the colour(s) of the letters and their display over the page/screen. Specifically, handwriting makes meaning through the type of calligraphy, the writing materials (e.g., pen, lipstick, pencil or marker), and the writing supports (e.g., paper, palm of the hand, duct tape or toilet paper)<sup>45</sup>, while typewriting makes meaning through the type of fonts, their size and, in dynamic texts, their appearance through time (e.g., whether typing appears all at once, letter by letter, word by word; whether it scrolls on the screen and from which direction<sup>46</sup>).

Although many of these features are usually not considered as significant for a traditional linguistic analysis, they are highly significant for the meaning that is made in videos. So, for example, the toilet paper used as a support for writing in the video of Fig. 6 contributes both to the humorous meaning of the video and to differentiate its representation from others’ (i.e., showing originality).

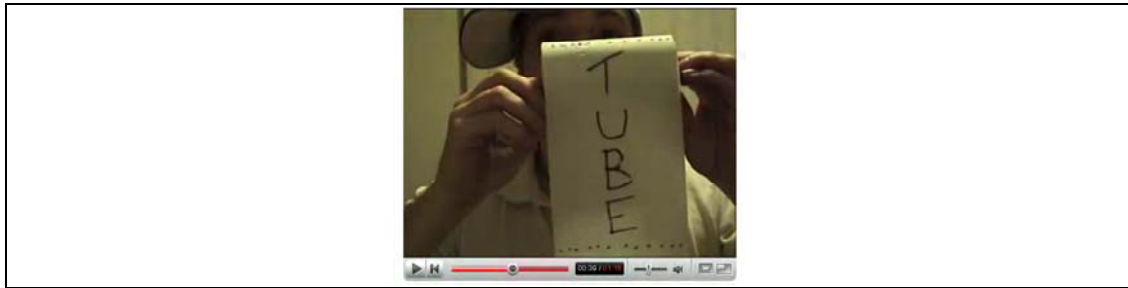
Indeed, every given resource – no matter its level or system –makes meaning if it is made salient by its producer and if it is perceived as salient by the viewer, so that multiple levels and systems of semiotic resources can be distinguished for each and all modes employed in videos, according to the level of delicacy and focus of the analysis (i.e., according to the researcher’s interest).

---

<sup>44</sup> In Saussure’s *Cours*, the term was ‘associative’; in the linguistic tradition it has been turned into ‘paradigmatic’ because of Saussure’s discussion of inflectional paradigms as examples of this axis.

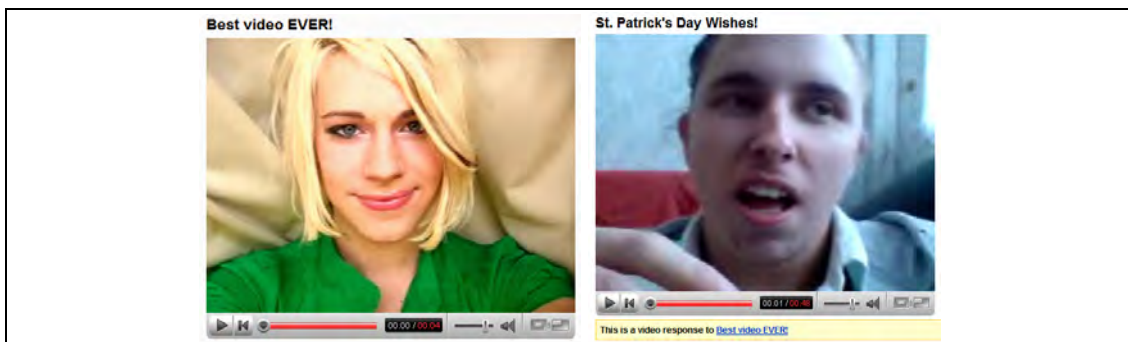
<sup>45</sup> All these instances of writing supports and materials are attested in the videos of the corpus.

<sup>46</sup> The scrolling of typewriting on the screen from bottom to top, for example, is the signifier of a filmic text type, i.e., the closing credits.



**Fig. 6 Toilet paper used as writing support.**

Moreover, as will be seen in detail when analysing the threads in the data, what is made salient by the producer of the text can be disregarded by the viewer who can choose to take up and make salient other elements. So, for example, as shown in Fig. 7 (and further analysed in Chapter 6, Section 3.5.1), the video titled ‘Best video EVER!’, which features a (You)Tuber’s face blinking and smiling at the camera, is responded by a video titled ‘St. Patrick’s day Wishes!’, in which a videoblogger wishes his viewers a good St. Patrick’s day. This way the respondent makes salient the initiator’s emerald-green colour of the blouse (rather than – say – the blinking) and relates it to the day when the initial video was posted (17 March, St. Patrick’s day, celebrated in the Irish calendar). In other words, in interactional exchanges, not only does the sign-maker select a mode and a form to produce a sign and, eventually, make it salient, but also the respondent chooses to take up a prompt, out of the many possible ones, and makes it salient in her response, according to her interests in responding.



**Fig. 7 The ‘Best video EVER!’ initial video and the ‘St. Patrick’s Day Wishes!’ response.**

Besides the selection among modes and modes sub-features, multimodality in videos involves meaning which is given by the intertwining of resources made through a multiplicity of modes, deployed both in time and space. So for example, the snapshot which in Fig. 4 was used as exemplifying drawing, makes its humorous meaning by the relation between the drawing (i.e., the cloud sketched on a post-it) and the handwriting (i.e., the country, UK, whose rainy weather is a notorious topic in jokes). Analogously, the snapshot referred to music in Fig. 4 actually represents a handwritten paper which draws the viewer’s attention to the soundtrack playing in the video and thus makes it salient.

In other words, each sign, apart from being the fruit of a paradigmatic selection, makes meaning by inserting itself in a 'syntagmatic' (Saussure, 1931) chain of resources; which resource is in relation to which other one depends on the relations made salient by the sign-maker (i.e., the producer) and on the relations perceived as salient by the meaning-maker (i.e., the viewer).

All forms of communication make meaning through paradigmatic selections and syntagmatic relations. Furthermore, this wide availability of (selection among, and combination of) resources is not unique to video-interaction; yet the particular range of resources is. So, for example, face-to-face encounters rely on a wide range of modes to make meaning in three-dimensional – rather than two-dimensional, as in videos – space and in time. This range includes body contact and proximity, which, in turn are excluded among video-interactants. Conversely, video-interactants can employ some resources which are not generally available in face-to-face encounters, such as filmic effects or computer-generated resources (e.g., typing on the screen, animations etc.).

Videos are dynamic texts, so that syntagms develop through time, and this differentiates video-interaction from other types of visual communication (like emailing or drawing, for example, whose syntagms develop in space). Besides, videos make meaning also by the deployment of resources in space, differently from types of auditory communication, e.g., phone conversations or recorded music (which only develop through time).

In other words, video-interaction can rely on both visual and auditory perception in simultaneity and through time. This assimilates it to face-to-face encounters and to live concerts or dance, while it distinguishes it from interactions which rely (mainly) on one of the two perceptive sources, either the visual (emailing, chats, various forms of visual art, architecture, dress etc.) or the auditory (phone, recorded music playing etc.). In turn, by relying exclusively (and simultaneously) on the visual and auditory, video-interaction does not afford other types of perceptions, like touch, for example, or smell and taste, which can all be used in face-to-face encounters and can be also predominant in specific types of communication (e.g. wine-tasting, for example).

The intertwining of paradigmatic selections and syntagmatic relations among semiotic resources in a video become even more complex in the interaction among videos. Indeed, when videos respond to one another, the resources selected and used in a video are interrelated (by difference and similarity) to those of the video they respond to. Again, this multimodal complexity is by no means rare or exceptional in human communication and interaction. For example, in face-to-face communication meaning is made by many of these modes, both in space and time; so, in greetings, one can decide to use spoken language (e.g., 'hallo'), together with a facial expression (e.g., a smile), gaze and body posture (both directed to the greeted person). In turn, the greeted person can (cor)respond with the same multimodal deployment or can choose – for various reasons – to produce variation (e.g., by

avoiding gaze contact or by avoiding speech and, in its stead, employing gesture, like a wave of the hand). Not only the meaning of a wave of the hand can be differently perceived from a spoken 'hello', but, in interaction, the selection of the former to respond to the latter (or vice versa) can be (interpreted as) significant. An avoided gaze contact surely is, although this sign may be assigned a different meaning in different cultures and situations.

## **1.2 Homogeneity and bidirectionality**

As introduced earlier (cf. Chapter 2, Section 4.1.1), video-interaction is characterized by a response to a prompt which is made by means of the same type of text (i.e., a video which responds to a video); hence it is here called a 'homogenous' type of interaction, distinguished, for example, from the interaction which takes place on the Website by means of two different texts, e.g., as with written text commenting on a video.

Given that the second video is linked as a response to the prompting one, video-interaction is bidirectional – one video addresses the other – and, thus, it is characterized by homogenous *exchanges* (cf. Chapter 2, Section 4.1.1). This latter feature differentiates video responses from the so-called 'related videos' on the Website, for example, which are displayed on a video page only by virtue of similar tagging but which are not addressed one to the other (cf. the discussion in Section 2.2). Bidirectionality and homogeneity differentiate video-interaction also from other video-based forms of communication, like TV broadcast, which is unidirectional (a 'one-to-many' form of communication) and which cannot generate homogeneous exchanges. Indeed the response to a TV programme cannot be the viewers' broadcast of a 'TV programme response' and, in order to interact with the TV broadcast, viewers can respond only with acts which are different from the prompting one, e.g., by continuing watching, by switching off the TV, by changing the channel, by writing an email to the programme editors etc.

Conversely, homogeneity and bidirectionality are features shared by, e.g., email, chat and text messaging exchanges, forum discussions, dance, music and singing improvisations, online gaming, face-to-face and phone conversations.

## **1.3 Publicity**

Bidirectionality combines in video-interaction with publicity. All video-exchanges can be publicly monitored, unless the videos are set to private by their uploaders. This entails that, apart from the participants, a third (multiple and unknown) entity, the viewers, is potentially involved in the interaction. Publicity is relevant to video-interaction in that, on the one hand, viewers can decide to participate and join the interaction (by posting a response in their turn), so that the exchange can become multiple; on the other hand, participants shape their texts while being aware of a

public possibly – and often hopefully – viewing them, and, as a consequence, signs in a response can be meant– both intended and interpreted – to address either the other interactant or the viewers (or both).

When posting a video, the initiator can address the public in many means, e.g., with variously produced greetings and by gazing at the camera. In contrast, by – say – gazing at the camera, the respondent addresses the initiator (*and* the public), thus establishing what in filmic conventions is called a pair of ‘reverse angle shots’ (Kress and van Leeuwen, 1996, 2006: 259). This way, the initiator’s and the respondent’s gaze constitute two ‘disconnected syntagms’ (Kress and van Leeuwen, 1996, 2006: 259), displaying a ‘disconnection of reactor and phenomenon’ (Kress and van Leeuwen, 1996, 2006: 259), which, in films ‘have to be “matched” carefully, to restore the connection’ (Kress and van Leeuwen, 1996, 2006: 259).

In video-interaction these signs in the two videos are connected by the viewer as if distance of time and space were not there; even more, they do not need to ‘match’ as in a film (by e.g., means of the same background, lightning effects, camera angle etc.), because it is understood that the exchange is asynchronous and not in spatial co-presence (see Sections 1.4 and 1.7). In other words, when viewers know that a video is a response to another one, they automatically connect the two disconnected syntagms and feel to be secondary-addressee (this is not possible in public discussions; one either looks at the person she responds to or looks at the whole audience). This convention of interpreting unmatched syntagms as connected is practiced also within single online videos, e.g., in the case of mashup videos, which are made by an assemblage of shots of different videos (cf. for example the video response described in Chapter 6, Section 3.3.2c, in which a fictional dialogue between the (You)Tuber and the thread initiator is represented by interposing shots of the (You)Tuber and shots selected by various videos of the initiator)

As discussed in Lange (2008), participants on (You)Tube exploit various resources to modulate the extent of publicity of their videos, which range from using the related option on the interface (the ‘set to private’ option), to using the video paratext (by using more or less popular tags and thus making the video more or less retrievable on the Website), as well as from the type and amount of information disclosed in the video itself to the one disclosed in the uploader’s profile (so as to reveal or disguise the identities of the protagonists and authors of the video).

When this is done in interaction, the extent of publicity of a video does not depend only on the one selected and modulated by its uploader; indeed, given that video-interaction is an exchange composed of (at least) two videos publicly linked one to another, the publicity of one is also the result of the practices used by the other video’s interactant. This works similarly in other forms of communication. Indeed, usually, either interactants agree on the extent of publicity of their interaction or the latter is given by the result of conflicting behaviours (e.g., an interlocutor revealing to a third party the content of the interaction makes it *ipso facto* more public,



independently of the other interlocutor's agreement). Of course, this works both ways. So, the extent of publicity of one's contribution can be potentially increased thanks to the other interlocutor's rate of 'exposure'. In this case, posting a response to a very popular video is not far from intervening at a VIP's conference (although viewers of a VIP's conference usually cannot avoid to attend all interventions, while in video-interaction each response has to be clicked purposely by the viewer).

#### **1.4 Asynchronicity**

Video-interaction exchanges are not conducted 'live'. The time of production and the time of reception of a video do not coincide. On the one hand, unlike in face-to-face interaction, this prevents the producer to get live feedback from her interlocutor so that she cannot change the production live on the basis of the viewer's reactions; it also prevents the producer from being interrupted, so that interpositions and overlapping never occur. On the other hand, asynchronicity allows sign-making to be planned, so that a video can be designed (through a script) and edited at will before uploading the version which is considered as the most apt to its uploader's interests.

At the same time, at the level of reception, asynchronicity allows a video to be (re)played at will (or stopped) by the viewer so that the 'control' of the production (the possibility of planning, i.e., design) finds its analogue in the control of the reception (the possibility of checking back the whole video or some of its details and to (re)elaborate the interpretation). This differentiates video-interaction from real-time types of interaction, like concerts, online games or face-to-face encounters, and from video-conferencing (the other type of interaction which employs videos), while it assimilates it to, e.g., epistolary forms of interactions, recorded and text messaging.

#### **1.5 Disembodiment**

Video-interaction is made of disembodied texts (which, as seen earlier, can represent both embodied and disembodied signs); they are tangible entities (made of pixels and bytes). This means that their products are separated from their producers and can be replayed at will by viewers after the producers have produced them. This assimilates online videos to most forms of written texts, to architecture, to recorded sounds and music, to dress, to online games etc. However, differently from other forms of disembodied communication, videos enable embodied resources to be enacted (in front of the camera) in disembodied texts. Hence, although asynchronous, video-interaction can be literally 'face-to-face'; indeed, videos can show what is presented as its producer's face looking at the camera, i.e. at her interlocutors (cf. Chapter 2, Section 4.4).

This combination of embodied signs in disembodied texts differentiates video-interaction from most forms of asynchronous interaction. So, interestingly enough, in showing the face of the interactant, video-interaction overcomes some of the

restrictions of written computer-mediated communication, in that body and facial expression, context and spoken language can all be deployed in videos. Yet the study in Adami (2009a) reveals that the typical computer-mediated communication ‘paralinguistic’ signs, the so-called ‘emoticons’ (Crystal, 2001), are still very much used in videos. On the one hand, thus, this signals that these icons are widening their communicative functions, beyond being a mere support to what the written mode lacks, in terms of intonation and body expression. On the other hand, and maybe even more interestingly, this proves that multimodal redundancy plays an active role in communication and is never superfluous<sup>47</sup>. Certainly, disembodiment and asynchronicity may favour redundancy, since the producer cannot have immediate feedback and thus monitor the receiver’s attention, understanding and/or agreement.

## **1.6 Online**

Video-interaction employs a digital technology and occurs online. Online interaction does not require geographical proximity of the interactants, while it requires hardware equipment and Internet connection to take place (and both are not accessible to a large part of the world population). Furthermore it entails a certain ephemerality of the exchange, which can be modified or cancelled by the interactants or by the Website owners (on the server) after it has taken place (for the characteristics of online communication, cf. Herring, 2001). Rather than on the basis of geographical proximity, interactants can meet in this semiotic space on the basis of interests and thus they can negotiate their own sub- or micro-cultures, practices and conventions not only within a shared (e.g., national) culture but also beyond their cultural differences. In this regard, video-interaction shares all features of online communication, whose literature is extensive (cf., among others, Barnes, 2003; Baym, 2000; Bell, 2001; Gross and Acquisti, 2005; Hamman, 2001; Herring, 1996, 2001; Holt, 2004; Jones, 1998; Kendall, 2002; Lewis, 2007; Melican and Faulkner, 2007; Sandvig, 2006; Sheehan, 2002; Shepard and Watters, 1998; Thurlow et al., 2004; Weiss, 2007; Wellman and Gulia, 1999; Xia et al., 2007; Yahia et al., 2007).

Here again, the fact that video-interaction enables embodied modes to be deployed online differentiates it from other forms of online communication, in that e.g., cultural differences can be expressed and specific conventions can be negotiated also for modes which are normally ruled out online, such as gestures, facial expressions and body posture (not to mention spoken language).

## **1.7 Distance**

Like telephone conversations or epistolary exchanges, video-interaction is a distant form of communication, in that it does not require the spatial co-presence of its

---

<sup>47</sup> In this view, the conception that anything is ‘para’ to language is definitely misleading and misrepresents the complex intertwining of functions played by each mode in producing meaning.

participants. This prevents body contact among interactants and the use of the system of proxemics as a resource to make interpersonal meaning. Conventionally, proxemics is used to establish power relationships<sup>48</sup> between the interactants. These need to be expressed by other resources in video-interaction, such as the camera angle and the type of shots, i.e., the distance and the horizontal and vertical orientation of the represented participant in relation to the camera, as in traditional filmic and imaging conventions (for a detailed discussion, cf. Kress and van Leeuwen, 1996, 2006: 114-153).

## 1.8 Multiple mediation

Kress and van Leeuwen (2001) distinguish ‘production’, ‘design’, and ‘distribution’ as the three processes of semiosis, i.e., the processes through which meaning is made and discourses are realized. These processes are specifically assigned to specific media and specific agents in video-interaction.

The mediation involved in video-interaction is multiple. In the first place are all the media used to produce the representations in a video; these can include the body or some bodily parts, as well as writing supports and materials, the media which produce sounds and the hardware and software tools which produce computer-generated resources. A second-layer of meaning is produced by the (Web)camera, which films and edits what is enacted in front of it, and the hardware and software tools for image creation and editing. All these media are in charge of the process of text ‘production’; the differentiated availability of these media determines different text productions and, therefore, different meanings.

Sometimes, before producing their videos, (You)Tubers create scripts and storyboards for their videos. The media – such as paper – involved in this process of semiosis are in charge of the ‘design’ of the text. The process of design involves also the selection of existing materials, such as TV excerpts or other *YouTube* videos.

The text is then uploaded on *YouTube* Website, the online medium, which is in charge of its distribution. The distribution process adds further meaning to the text, in terms of various effects which are due to structural restrictions (e.g., the fact that videos on *YouTube* can last max. 10 min.) and to visual conditions (e.g., the overall layout of the page where the video is displayed), just to mention a few. The interface – and their owners – greatly influence the meaning that is made in videos, as discussed in the next section. But video makers can also act on the distribution medium, by modulating, for example, the extent of publicity of their videos (as discussed earlier in Section 1.3). In other words, video-interaction, unlike other forms of communication, has a high convergence of the various semiotic processes and their media in the hands of the author of the text. This differentiates online video

---

<sup>48</sup> For proxemics as a resource for power, cf. Hodge and Kress (1988: 52-59).

sharing from filmic production (in which the design, the production and the distribution of the text are assigned to different and specialized entities, i.e., the screenplay writer, the director, the actors, the cameraman, the special effects editors, and the distribution corporations) for example or from communication in the context of education, which is currently undergoing a process of specialization (as discussed in Kress and van Leeuwen, 2001: 122-124).

## 1.9 Corporate interface distribution

Video-interaction occurs by means of an interface (a Website) which is owned by a (billionaire) Corporation, i.e., *Google Inc.* They bought *YouTube* in November 2006 for 1.65 billion US\$ (in *Google* stocks) from their inventors, Chad Hurley, Steve Chen and Jawed Karim (who had launched it online in February 2005). The interface has technical affordances, which are examined in detail in the next section of the chapter, while the owners have their own policy and (commercial) interests.

*YouTube* policy is published on the site. So, for example, the ‘Terms of Use’ page<sup>49</sup> specifies that, although the uploaders are the owners of their videos, the Website owners reserve the right of using any of the materials which are published on the Website. Moreover they prohibit (You)Tubers to use the Website for commercial purposes, while these are well pursued by the Website owners, who host advertisements and corporate channels on *YouTube*, by virtue of commercial agreements with various corporations<sup>50</sup>. The Website owners prohibit also the use of unauthorized copyright material in videos, which may result in the removal of the video and, in case of repeated violations, in the termination of its uploader’s account.

The Website owners set also restrictions to contents. So, the ‘Community Guidelines’<sup>51</sup> express the owner’s prohibition of ‘pornography or sexually explicit content’, of ‘bad stuff like animal abuse, drug abuse, under-age drinking and smoking, or bomb making’, of ‘[g]raphic or gratuitous violence’, including videos representing ‘someone being physically hurt, attacked, or humiliated’, and of ‘things intended to shock or disgust’ such as ‘gross-out videos of accidents, dead bodies or similar’. In spite of these restrictions, *YouTube* proudly affirms that ‘freedom of expression’ is granted; nevertheless ‘hate speech’ is not allowed, which is defined as

speech which attacks or demeans a group based on race or ethnic origin, religion, disability, gender, age, veteran status, and sexual orientation/gender identity

Also prohibited are certain types of behaviour, such as

---

<sup>49</sup> <http://www.youtube.com/t/terms> (Retrieved 8 November 2008).

<sup>50</sup> Cf. Burgess and Green (2008) for a discussion on the conflicts arising between corporate commercial interests and (You)Tubers’ participatory needs.

<sup>51</sup> [http://www.youtube.com/t/community\\_guidelines](http://www.youtube.com/t/community_guidelines) (Retrieved 8 November 2008).

predatory behavior, stalking, threats, harassment, intimidation, invading privacy, revealing other people's personal information, and inciting others to commit violent acts or to violate the Terms of Use

Finally, since the owners assume that '[e]veryone hates spam', they do not permit to 'create misleading descriptions, tags, titles or thumbnails in order to increase views' and to 'post large amounts of untargeted, unwanted or repetitive content, including comments and private messages'.

The Website owners endeavour to reconcile the prohibitions they need to impose to prevent themselves legal inconveniences (and image drawbacks in the eyes of mass media and public opinion) and the image they want to give of themselves to their 'users' and the participation they need to have by them (as will be seen, also 'users', in their turn, endeavour to weave these constraints, by using the affordances of the interface creatively according to their purposes). In this sense, apart from listing what is prohibited, the guidelines encourage users to 'dig in and get involved':

Remember that this is your community! Each and every user of YouTube makes the site what it is, so don't be afraid to dig in and get involved!

Together with that, they invite people to provide content, to give feedback and to contribute to the Website policing activity.

*YouTube* policy is also given by the kind of (young, democratic and groundbreaking) image that *Google* corporation wants to give about themselves and *YouTube* Website, so, every constraint or prohibition (i.e., censorship) is communicated in a 'cool' way, by hedging it, and by supporting it by means of discourses on the community, as the introduction to the 'Community Guidelines' exemplifies:

We're not asking for the kind of respect reserved for nuns, the elderly, and brain surgeons. We mean don't abuse the site. Every cool new community feature on YouTube involves a certain level of trust. We trust you to be responsible, and millions of users respect that trust. Please be one of them.

([http://www.youtube.com/t/community\\_guidelines](http://www.youtube.com/t/community_guidelines) Retrieved 8 November 2008)

Of course, what the owners foster and what they prohibit on the Website is not only expressed verbally in the pages devoted to the Website policy. It is also – and maybe to a greater extent – 'embodied' in the affordances of the interface, i.e., what is materially possible or impeded to do on the Website, together with what is promoted (i.e., more immediately available) and what is disfavoured, what is socially prized and what is penalized by means of the interface options. So, the Website slogan 'Broadcast Yourself' fosters individual protagonism and visibility (rather than, say, action and participation); the Website homepage<sup>52</sup> fosters 'popularity' and 'common practices': the first section immediately below the Website masthead is 'videos being

---

<sup>52</sup> <http://www.youtube.com/> (Retrieved 8 November 2008).

watched right now’<sup>53</sup>, while, below that, a selection of featured videos appears (selected by the *YouTube* editors), which fosters ‘competition’ in achieving visibility. The Website page dedicated to videos<sup>54</sup> fosters again popularity, competition and visibility, by listing several top charts which display the ‘most’ (viewed, rated, discussed, responded, commented etc.) videos according to various periods of time (‘today’, ‘this week’, ‘this month’ and ‘of all time’). Analogously, each video page shows statistics in terms of views, rating, bookmarking etc. These are choices which are made by the interface designers according to their interests and which foster certain practices (without although necessarily impeding others) and goals. For example, the ‘most’ top charts do not have their counterpart homologues in the ‘least’ ones; this promotes visibility and popularity while disfavours ‘rarity’ (so that there is no means of browsing among videos which have never been watched by anyone, for example). As a result, participation is connoted by majority values (i.e., quantity) and so is content quality, in that, what is viewed by most people is implicitly communicated as what is worth most viewing.

### **1.10 The place of video-interaction in the semiotic landscape**

To sum up the discussion concerning the structural characteristics of video-interaction, Fig. 3 represents its distinctive features in comparison with other forms of communication. Of course, the list of forms of communication is not exhaustive and could be enlarged at will, but it can usefully attempt a first mapping of the place of video-interaction in the contemporary semiotic landscape.

---

<sup>53</sup> From time to time the Website owners change the layout of the interface. Changes have been monitored since the beginning of the corpus collection (August 2007) up to September 2008. Hence, the final layout referred to in the present work is the one in force at the time of the end of the monitoring period (end September 2008). How the interface is (slowly) evolving trying to satisfy the participatory interests of (You)Tubers, cf. Burgess and Green (2008).

<sup>54</sup> <http://www.youtube.com/browse?s=mp> (Retrieved 8 November 2008).

	Video- interaction	Face to face interaction	Email exchanges	Collective art improvisation	Phone exchanges	TV broadcast	Video- conferencing	Chat	Online video- gaming
embodied and disembodied multimodality	+	-	-	+	-	+	+	-	+
homogeneity and bidirectionality	+	+	+	+	+	-	+	+	+
publicity	+	+	-	+	-	+	-	-	+
asynchronicity	+	-	+	-	-	+	-	-	-
disembodiment	+	-	+	-	-	+	-	+	+
online	+	-	+	-	-	-	+	+	+
distance	+	-	+	-	+	+	+	+	+
multiple mediation	+	-	-	+	-	+	+	-	-
corporate interface distribution	+	-	+	-	+	+	+	+	+

**Fig. 8 The distribution of distinctive features of video-interaction compared to other forms of interactions.**

## 2 THE ‘VIDEO RESPONSE’ OPTION: AFFORDANCES AND PRACTICES

As often happens in human interaction, the distinctive characteristics detailed in the preceding section are helpful to understand the possible ‘mechanisms’ of video-interaction; however these features only set the structural ‘ground’ of what is possible to do in video-interaction (i.e., its affordances) and do not tell much about how video-interaction is actually practiced. In other words, what discussed above is only what is structurally favoured by the interface affordances, while participants may adopt various practices according to different interests and different ethics<sup>55</sup>. The complex intertwining between affordances, practices and policies and its effects on video-interaction is examined in the present section.

<sup>55</sup> Intended here as ‘evaluations on aesthetics’, cf. the definitions of ‘style’, ‘aesthetics’ and ‘ethics’ in Kress (2008).

## 2.1 The Video Response Option

On May, 16 2006 the *YouTube* editors announced the introduction of the video response functionality by posting the following blog entry on their blog:

### Video Responses

We recently noticed that within many of the different ecosystems on YouTube our users are doing something really cool - they're communicating with each other through their videos. Text comments and messages are great, but our users have once again created something really innovative completely on their own - video responses. It's been amazing to watch our users create an entirely new mechanism for communicating with one another. However, one of the challenges with these video dialogues has been there is no way to 'link' your response back to the original video. To encourage and simplify this type of communication we just launched a new Video Response feature that will allow you to upload your own video reply while you're watching a video. Just look for the 'post a video response' link on any video watch page. All video responses will show up directly beneath the original video (just like text comments).

Check it out and let us know what you think!

Maryrose

The YouTube Team

(<http://www.youtube.com/blog?entry=-UV0BmDAq1c> Retrieved May, 26 2008)

Following the Website editors' account, the video response functionality has been added as a consequence of a practice that had already taken place on the Website. The post highlights the novelty of this communication practice – named 'video dialogue' – as well as the function of the new functionality of facilitating it.

Therefore, a semiotic practice among (You)Tubers, who used the existing affordances of the interface (the related videos functionality) to their own purposes, has been taken up by the Website owners, who have changed the interface accordingly. In multimodal discourse analysis terms (Kress and van Leeuwen, 2001), the producers' innovative use of the medium affordances according to their interests has led the designers to change the affordances of the distribution medium, i.e., the Website interface. In even more general semiotic terms, a socially shared semiotic practice has affected the semiotic structure (as much as the Saussurean *parole*, acknowledged socially, has entered the *langue*). In its turn, this change in the structure, as will be seen, is giving rise to further innovative practices.

## 2.2 Related videos vs. video responses

The affordances of the 'related videos' functionality and of the 'video responses' option are different.

Related videos are automatically displayed on the right of a given video on the basis of linguistic relations of occurrence, in much the same way as search engines work. This means that videos are shown as 'related' on the interface by virtue of identical



words occurring in the video title, description and keywords given by the video uploader, i.e., through tagging, (the categorization by means of user-generated tagging is called ‘folksonomy’, a blend between ‘folk’ and ‘taxonomy’, cf. Cattuto et al., 2007). Therefore relatedness among videos is established through intertextuality, by means of lexical occurrence in the paratextual information of the video page.

Hence, prior to the introduction of the video response option, (You)Tubers who wanted to relate their video to another one (in terms of, e.g., a reply to its content, a remix, a parody or a quotation) could only use the existing possibility to establish a visible ‘link’ between the two videos, i.e., they could exploit the related video functionality by using the same words of the original video in their video title, description and tags. In this way their video thumbnail would be displayed – together with many others which shared the same keywords – in the ‘related videos’ section on the right of the original video homepage (see Fig. 9); in turn, the original video thumbnail would be displayed among the related ones of their own. However, as the Website admits, the search algorithm which generates related videos is ‘mysterious’<sup>56</sup>, hence one could never know ‘how much related’ her video would result and, consequently, where it would appear among the many related videos displayed at the right of the original video. Indeed, the ‘related video’ section is a scroll-down window which readily displays on the video page only the first five thumbnails (see Fig. 9); other 15 thumbnails appear in succession by scrolling down the bar on the right of the window, and, at the bottom of these first 20 thumbnails, a link directs to ‘see all [N.] related videos’. Therefore the affordances of the layout of the related videos section make the first five more related videos immediately retrievable on the video page; some effort (i.e., scrolling down the bar) is needed to retrieve the other 15 more related videos, and even more effort (i.e., scrolling down the bar + clicking on the ‘see all related videos’ link) is needed to retrieve the related videos ranking from 21<sup>st</sup> on. Understandably, the higher the relatedness score of a video (‘relevance’<sup>57</sup>, in Information Retrieval terms), the more readily its thumbnail is retrievable in the ‘related video’ section. Yet, because the score is calculated by the interface algorithm, (You)Tubers have no full control on the extent of relatedness of their video with the original one.

In turn, the ‘video response’ option links a video to another one independently from their intertextual relatedness in the linguistic paratext. The link is created by the response uploader and can be confirmed or refuted by the initial video uploader. Hence, the video response option is more oriented towards the agency of the participants in the interaction, who intentionally create a link among their videos, rather than relating them by linguistic means in the paratext.

In sum, while the related videos section relies on inter(para)textuality, and may be

---

<sup>56</sup> <http://help.youtube.com/support/youtube/bin/answer.py?answer=70181&topic=10507> (Retrieved June 2 2008).

<sup>57</sup> Here the term ‘relevance’ is avoided because of its possible confusion with Sperber and Wilson’s Relevance Principle (cf. Chapter 2, Section 1.3).

exploited for interaction by (You)Tubers, who however do not control the result, the video response option relies directly on the interaction between (You)Tubers, who have full control of the link between their videos.

The introduction of the video response functionality has not replaced the ‘related video’ one, so that now (You)Tubers have two different means of linking videos among them and they can use them simultaneously. In this way, any video can be a response to a given one and, at the same time, can appear as related on the right of another one, by means of similar words occurrence in the title, description and tagging. In the corpus of this study, for example, the initial ‘@---Where do YouTube---@’ video is a response to an apparently totally unrelated one (‘≈Off The Hinges≈ 2 (rotoscope, magic & music)’), while in its related videos section are videos with the word ‘youtube’ in their title/description/tagging.

This example testifies to the fact that neither the related video functionality nor the response functionality prevent topic unrelated videos (i.e., ‘globally incoherent’, in van Dijk’s terms; cf. Chapter 2, Section 2) to be linked one to another; indeed, by relying on the paratext for the former and on the (You)Tubers’ agency for the latter, none of the two functionalities links videos on the basis of what is represented in them. Therefore they both can be used by (You)Tubers for their purposes, which may well go beyond relatedness and content-related response (as in the case of what is traditionally considered as ‘spam’). Fig. 9 indicates the related videos section, the video responses section and the response link as displayed on a video page.



Fig. 9 The related videos section, the video responses section and the response link on a video page.

The (a) intended purposes of the functionality, i.e., the primary use for which its designers made it, (b) its affordances, i.e., what it enables (You)Tubers to do, and (c)

its effective uses, i.e., what (You)Tubers do with it according to their purposes, are three distinct and interrelated factors which always come into play in video-interaction; each influences the others to a certain extent, but none of them determines the others in a linear way.

The next sections illustrates what it is technically possible to do through the video response option, before analysing the semiotic affordances which derive from these technical possibilities.

## 2.3 Technical affordances

The Website section ‘YouTube Glossary’ defines a video response as follows:

Video Response - A video can be associated as a reply to another, much like a text comment.  
(<http://help.youtube.com/support/youtube/bin/answer.py?answer=70181&topic=10507> Retrieved May 26 2008).

Thus, according to the Website designers, the option creates a ‘reply’ by association. The video uploader can (a) disable responses, so that no video can respond to hers, (b) allow video responses to be automatically posted to her video, without any prior approval, or (c) choose to be asked for approval for any of them to be displayed as video responses on the video page. Each uploaded video file cannot respond to more than one video, while every video response can receive responses in its turn, so that further levels in the thread can be created.

When a (You)Tuber clicks the ‘Post a video response’ option, the following message appears behind the video thumbnail:

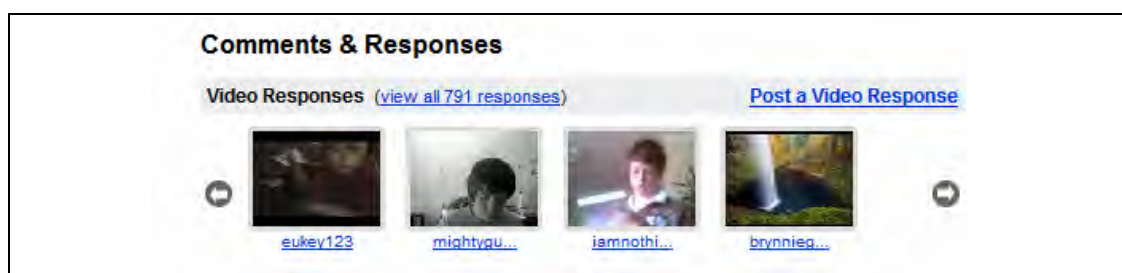
Ever wanted to talk back to a video? Now's your chance—you can upload a response to this video and we will link them together. You can record a new video, choose from the videos you already have, or create and upload a new video.

Therefore, a (You)Tuber can respond to a given video in three ways:

- by recording and uploading ‘live’ a new video, which is then created from scratch as a response. In this case, analogously to an email reply, the video title will automatically take the form of ‘RE: + [title of the initial video]’, while the video description will be ‘Video Cam Direct Upload’, thus informing on the ‘live’ recording of the video (both can however be modified by the uploader in a second time);
- by linking a video already uploaded in the past, so that the video response preserves its original title and description (which can be modified in a second time). In this case, given that a video can respond only to one video, the new response-link cancels any possible previous one with any other video;

- by making a new video ‘offline’ and uploading it as a response, after having filled all information such as title, video description and tagging.

On the video page, video responses are numbered according to their chronological order of posting and are displayed in reversed order (i.e., the most recent one is displayed first) from left to right, following Western reading conventions. As in the ‘related video’ section, only the first 4 thumbnails appear horizontally displayed in the video response section (at the bottom of the video on the video page), while two arrows – one on the left and one on the right of the section – allow the viewer to scroll up among the first (i.e., the most recent) 20 responses (Fig. 10).



**Fig. 10** The ‘Video Responses’ section displayed below the video on a video page.

In the video responses section, the link ‘view all [No.] responses’, directs to a video response page, where the initial video thumbnail is displayed on the top left, complete with its description, ‘statistics’ (no. of views, date of uploading, rating etc.) and username of its uploader (see Fig. 11). Below the initial video thumbnail, up to 60 responses are displayed on 10 rows of 6 thumbnails each; below each of them some information on the video file are displayed, namely its title, username, number of views, rating, video duration and number of response. If one follows the conventional Western reading path for Webpages (i.e., from left to right, from top to bottom) the responses thumbnails are displayed in chronological reversed order, from the latest to the oldest one (e.g., response no. 101 is on the left of response no. 100).

At the bottom of the page (Fig. 12), a numbered link directs to further response pages as well as indicating (by contrasting colours) the present one (in much the same way as the bottom link of search engines directing to other result pages). Since beginning 2008, a newly-introduced option enables all the responses to a given video to be played in succession, by means of a ‘Play all’ link in the response section on the video page (behind the ‘view all’ link). So, with just one click the whole video response thread can be played on the screen. Videos are played in succession from the most recent response to the oldest one, each video is displayed on its video page, so that all paratextual information is accessible, and a box on the top of the page informs ‘Now playing video responses to: [title of the initial video]’).

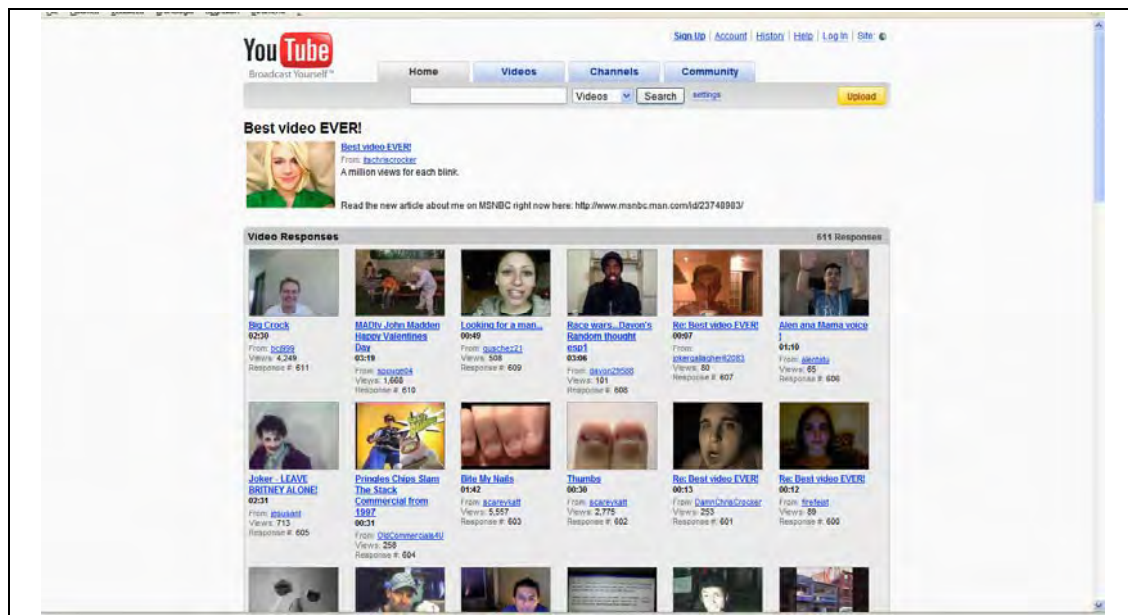


Fig. 11 The top of the video responses page.

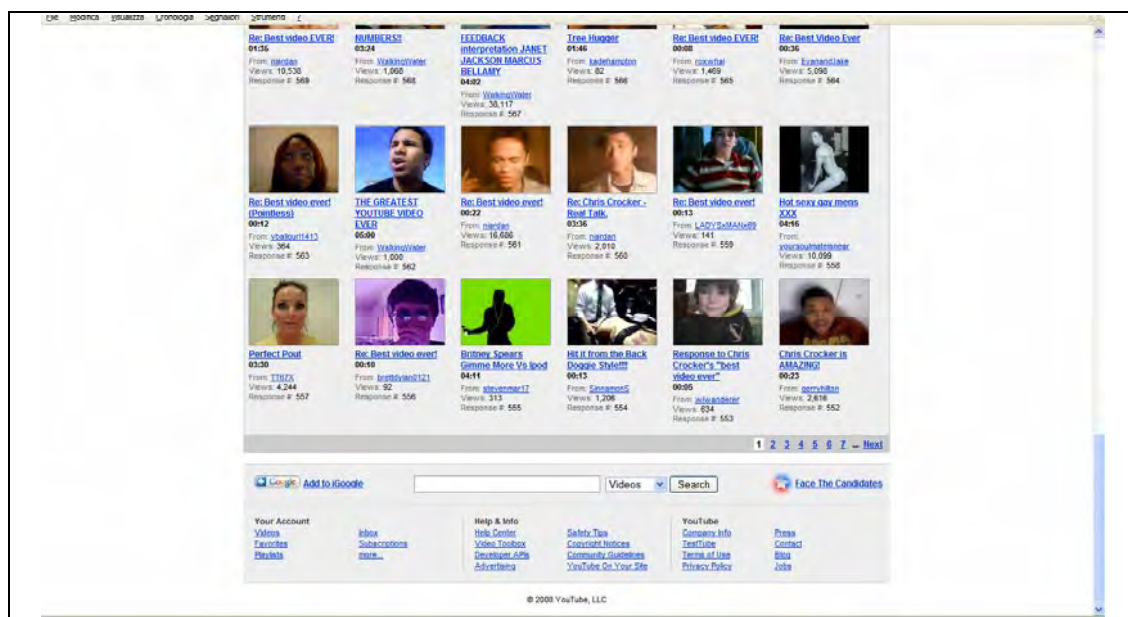


Fig. 12 The bottom of the video responses page.

## 2.4 The affordances in use

The above illustrated technical possibilities and constraints of the video response option generate semiotic affordances which are used according to the purposes of the interactants.

### 2.4.1 The approval/denial and the power of the initiator

By using the approval/denial option of video responses, the initiator can monitor the thread composition to a lesser or greater extent according to her interests and purposes. In other words, the (You)Tuber is the editor of the thread which is built by the responses to her video.

Hence, if the initial (You)Tuber is interested in receiving as many responses as possible, e.g., so as to get her video honoured among the Most Responded Videos top chart, she may accept all types of responses, even if topic-unrelated (this seems to be the case of ‘The Best Video EVER’ thread in our corpus; cf. Chapter 6). Conversely, if the (You)Tuber is rather interested in having a topically-coherent thread, she may filter out all responses that she considers unrelated; this is definitely the case, for example, of another Most Responded video, the ‘One World’ video, whose description reads as

Number One Most Responded Video of All Time. (*non-spam* responses that is)<sup>58</sup>

The practice of posting unrelated videos to the most responded ones is so frequent that a coherent composition of the thread achieves an added value, as evidenced by the above quoted description of the ‘One World’ video, which makes explicit that ‘spam’ responses are filtered out. At the same time, acknowledging this ‘spamming’ practice means also using it to one’s own purposes, as explicated in the title of another most responded video:

Everything (awaiting to be spammed)

Here again, an affordance of the structure (i.e., the fact that the response link is independent from topic-relatedness) is used by interactants according to their purposes, giving birth to differentiated practices (by the ones who welcome and solicit unrelated responses and by the ones who aim at giving extra value to their video-interaction thread in making explicit that it has been filtered from ‘spam’).

### 2.4.2 The denial and the process of interaction

Filtering out means censorship, in that every video uploader has the power to control all incoming connections to her video. To some extent in all forms of communication the interlocutor has the power to deny a response (e.g., in face-to-face conversations one can turn her back and go away from her interlocutor), however the effects of this censorship power are distinctively shaped in video-interaction.

On the one hand, the initiator’s denial of a video response affects only the *link* between the two videos, not the existence of the would-be response video, which

---

<sup>58</sup> My italics.

remains uploaded, online and accessible on *YouTube*. On the other hand, no trace of the attempted response link remains visible on the Website, so that only the initiator and the would-be respondent know the existence of this denied attempt to interact. This is usually not so in other forms of public interaction; indeed, either the denial prevents the response from taking place (not only its response status) or the unsuccessful attempt at interacting is public itself. Generally, the first case includes instances of asynchronous communication (e.g., forum discussions etc.), in which the moderator can filter out the responses (and, in so doing the message itself is erased, not only its response association); in the second case are instances of synchronous communication (both formal and informal), where the interlocutor's denial can prevent the response from taking place but its denial is monitored by the public.

Conversely, in video-interaction, what is publicly accessible is either (a) the result of a successful interaction (the initial video and the response) or (b) isolated instances of communication (two non-linked videos). Therefore all unsuccessful attempts of interaction are not visible.

As mentioned in Chapter 2 (Section 4.1.1), in terms of prompt-response, the semiotic acts involved in the exchange are:

1. the initial video (a prompt to what follows);
2. the attempted link by a second uploader of a video (a response to 1. and a prompt to what follows);
3. the initial uploader's acceptance or denial of the attempted link (a response to 2.).

The acceptance in 3. realizes the attempt in 2, so that on the interface a prompt (the initial video) is linked to a response (the video response). The denial results in two independent semiotic texts (the two videos), with no trace of any of the prompt-response relations that have occurred in the process.

In other words, what is observable on the Website are the products of interaction, not the process of interaction. In these products and in their relations are to be detected clues that reveal the process.

### **2.4.3 Responses creation and thread composition**

According to the way in which the response link is created, threads are composed by two main types of responses:

- the ones that are purposely created for that interaction (either recorded live or uploaded after making them offline);
- the ones which were previously created and uploaded and then are linked in a second time to a given video.

This not only differentiates video-interaction from other forms of communication but has also various consequences in the threads composition.

Firstly, threads are quite mobile in their composition; indeed, in order to get their video viewed, some (You)Tubers are likely to exploit all the possibilities given by the interface, including linking it as a video response to some other video; when another (more topic-related or more viewed or more recent) video strikes their attention, they may change the response link to this new one. For example, as shown in Fig. 9 above, in August 2007, the initial ‘Where Do YouTube?’ video was a response to a (topic-unrelated) video, titled ‘≈Off The Hinges≈ 2 (rotoscope, magic & music)’, authored by the same You(Tuber), *ChangeDaChannel*<sup>59</sup>. In beginning 2008, it moved its response link to another (topic-unrelated) one, titled ‘Britney Spears DILNO! Buffy’s a LESBO! J-NO LO!!’, authored by another popular *YouTube* profile, *WHATTHEBUCKSHOW*. As a consequence of this mobility of response links, the number of responses to a given video can not only increase but also decrease over time; this can seldom take place in other forms of interaction, while it does occur when we rearrange the folders and files of our pc (or the furniture of our rooms), for example, according to new interests and organizing criteria.

Secondly, responses are more or less topic-related according to whether they were created purposely to reply to that video or made in a previous time for other purposes and then linked to it, as discussed in the next section.

#### **2.4.4 Textual organization as clue for responses creation**

Although, apart from the date of the upload, there is no clue to tell videos created as responses from videos linked as responses in a second time, the Given/New textual organization in every response leads to discriminate between them.

Specifically, in threads started by videos requesting a topic-specific response, the ideational meaning of the responses in the thread is textually organized in two different ways, according to whether the response was created for the purpose of that interaction or if the response link was established with a previously existing videos. For example, the ‘Where Do YouTube’ thread (analysed in detail in Chapter 5) has two types of responses, which differ in their textual organization of the topic answer (i.e., on how the represented location – i.e., where the (You)Tuber (You)Tubes – is textually organized):

- responses where the location (i.e., the topic answer) is presented as focus/New information, either through the sequential organization (i.e.,

---

<sup>59</sup> Again, the video response option can be exploited for various purposes, such as linking the uploader’s whole video production so as to augment its visibility.



represented at the end of the video), or through the spatial organization (i.e., represented in the foreground). These are more likely to be videos that were created from scratch purposely to respond to the initial video, in that they have a textual organization which is unmarked for answering the question ‘Where Do YouTube?’, i.e., the location as focus/New, as in the verbal answer ‘I (You)Tube from London’ (Chapter 5, Section 3.3.2) .

- responses in which the location is presented as Given, either through the sequential organization (i.e., represented at the beginning of a video which then shows further details, as in the case of documentary-like videos about a town), or through the spatial organization of the video page (for cases where the location is in the video title and description, while the video shows some activities that take place or entities which are placed in the location), or as circumstantial element (e.g., the location mentioned at some point in the video, without being its topic). These videos were arguably created and uploaded for other purposes and then linked – and thus recontextualized – as responses to the ‘where do youtube’ video in a second time, since they have a marked textual organization for an answer to the question ‘Where Do YouTube?’, i.e., the location as Given (e.g., ‘Austin is the world capitol of live music’) or as a circumstantial element, as in the verbal answer ‘Coffee is delicious in Austria’ (Chapter 5, Section 3.3.3).

In sum, the fact that one can respond either with a previously-uploaded video or with a video purposely made for the exchange influences the composition of the thread in terms of different textual organizations of its responses. Indeed, the New/Given textual organization of the topic answer is such a clear-cut divide among responses that, when one knows how responses can be created, one cannot but relate the different textual organization to the two different possible ways of responding.

In other words, the technical affordances in the response option influence the different textual organizations of the responses in the thread, and, ultimately, the different degrees of topic-relatedness with the initial video. Even more, the use of this affordance in the (You)Tubers’ practice of interacting with videos is a factor which contributes to the acceptability of textual organizations of responses which would be conventionally considered as ‘marked’ in other forms of interaction.

This is therefore a first explanation of the reason why video-interaction employs patterns of relatedness and coherence which would not be considered as acceptable in other forms of communication. As a way of example, the following exchange is highly unlikely to take place in face-to-face conversation:

A: Where do you live?

B: Religion practices in Taiwan are pretty diversified...

In turn, the ‘Where Do YouTube?’ video-thread testifies to an analogous – yet successful – video exchange, when a response to the topic question is a documentary

titled 'Religion in Taiwan' which displays images of various religious practices and a voiceover introducing them verbally.

The spread of types of text productions which afford the selection and re-contextualization of existing materials (e.g., through 'copy-and-paste' or 'forward' processes) is likely to increase textual organizations of exchanges which would be traditionally considered as marked (for this affordance in mobile devices, cf. Kress and Adami, 2009). In turn, these new forms of representation are possibly influencing the textual organizations of old types of representations (so, for example, school teachers often lament the 'incoherence' of students' essays, made through a 'copy-and-paste' of resources selected from the Web).

#### **2.4.5 The 'Play all responses' option and the thread as an entity**

Before the introduction of the 'Play All responses' functionality, it took more effort to watch all video responses to a given video than to follow other reading paths. Therefore, it was very unlikely that all responses were watched in succession, and hence, what is here called a 'thread' was not an entity to viewers.

Indeed, from any video page, it took (1) one mouse-click to be directed to the video response page and watch the response. From there, it took at least two further mouse-clicks to (2) go back to the initial video (which starts playing again as soon as you are redirected to its page), and, once identified the successive video response in the sequence, to (3) go to its homepage and watch it. This way, watching each response involved three mouse-clicks and the annoyance of having to re-listen and re-watch at least some seconds of the first video every time before watching the following response (it is incredibly annoying to watch again even one second of a video that you have just watched). Alternatively, from the initial video page, you could (1) mouse-click the 'View all responses' link, which displays the page with all the responses thumbnails in sequence. From there, you could (2) click a response and watch it. Then, in order to see a second response, you had to (3) go back to the previous 'all responses' page and (4) click the second response and watch it. This second path avoided the replaying of the initial video but involved an additional mouse-click (four instead of three). The effort involved in the mouse-clicks added to the psychological 'failure' implied in the fact that in both cases the viewer had 'to go back' to an already visited page on the Web (similar to turning back the page in a book or walking back on a street up to a crossroad to take a different path). Other viewing paths were more readily available, such as 'hopping' from one video to its response and from there to its possible response or to one of its related video. Indeed, instead of the two clicks needed to go back and to watch another response, from any given video response it takes only one mouse-click to follow one of its related video (or response), which, furthermore, psychologically implies 'advancement', since the viewer never goes back to already visited content and keeps browsing new one.

Therefore, prior to the introduction of the ‘Play all responses’ option, viewers were more likely to follow related videos rather than watching all video responses, while only the participants in the interaction were likely to watch all the responses to a given video. Understandably, only the interactants’ interest was worth the effort needed to watch the whole thread; indeed, the initiator is generally highly motivated to watch the responses posted to her video (especially if she is planning to make a video-summary) and, presumably, the respondents are similarly motivated to watch the other responses in the thread, either to avoid repetitions before posting their own, or, after it, in order to see if someone has taken up their contribution.

In contrast, now that the whole thread can be played in succession, it is more readily accessible to viewers, since it takes little effort to watch it as a whole. Indeed, it takes only one mouse-click to play all the responses on the screen one after another and one needs neither to go back, nor to re-watch the initial video. A change in the medium affordances has made the thread much more an ‘entity’ also to viewers.

#### **2.4.6 The ‘Play all responses’ option and the sublevels in the thread**

With the introduction of the ‘Play all responses’ option, further sub-levels in the thread are more likely to be treated as separate entities from the initial video. Indeed, if a viewer chooses the ‘Play all responses’ reading path, video responses to these responses will not be played (what here is called sub-response is indeed displayed on the interface just as a response to its video). If the viewer chooses to follow one of these sub-responses, the ‘Play all responses’ interrupts and, in order to resume the thread, the viewer has to go back two pages and re-click the ‘Play all responses’ (which will start again from the most recent one).

Therefore, the video-thread as considered here (i.e., made of an initial video, all its responses, and all their sub-responses) is not an entity to the viewer. Either the reading path uses the ‘Play all responses’, and thus considers the first level of the thread as an entity, or it follows the ‘initial -> response -> sub-response’ links, as much in the same way as one follows the links to related videos.

However, in terms of the participants’ perspective, the data analysed here show that these second-level responses often still refer to the very initial video (e.g., second or third level responses still reply to the ‘Where Do YouTube?’ question, as discussed in Chapter 5 Section 4.1, or address the ‘Best Video EVER!’ initiator, as discussed in Chapter 6 Section 4). In turn, some comments on these sub-responses (e.g., ‘what was that?’, ‘I didn’t get it’, or ‘wtf?’ = ‘what the fuck?’) evidence to the fact that viewers still treat the response as only related to its immediate prompting video.

Therefore, participants and (non-participant) viewers have different perceptions of video-interactional exchanges.

#### **2.4.7 The sequential display and the values of ‘firstness’ and ‘newness’**

Due to the interface layout, which displays the video responses to a video by numbering them according to the time of posting, video-interaction is necessarily sequential. True, there might be overlapping when two videos are posted simultaneously, however, they are nonetheless displayed in a sequential order. Hence, although the authors who post their videos at the same time can not have previously watched each other’s video, any possible overlapping has no visible effects on the thread. This sequential feature creates first- and last-posted responses, so it creates first-time used semiotic resources and repeated ones (i.e., perceived by the viewers as ‘already done, nothing new’, if not even ‘copied from others’).

Being the first in doing something is highly valued, in Western society at large and on *YouTube* (which stems from it) in particular. This is attested on the Website by the very frequent advice for making good/interesting videos that (You)Tubers give in their ‘how-to’ videos (e.g., ‘do something that has never been done’). Moreover, although not overtly promoted, it is strongly fostered by the Website interface, whose top chart display system and arbitrary selection of featured video lead to equate ‘the most’ and ‘the best’ with ‘the first’ (cf. Section 1.9). ‘Firstness’ is inevitably associated with ‘innovative’ and therefore with the positive value of ‘creativity’. This is even more so when, as in video-sharing, the medium itself is relatively new and the product it delivers requires some innovative skills in mastering it.

Given that the sequential display of the responses is in reversed chronological order (i.e., from the most recent to the oldest ones, like the posts in a blog), the ‘first’ value is counterbalanced on *YouTube* by the ‘newest’ value, thus promoting fidelity in watching (the assumption is that old material is less interesting – and hence more difficultly recoverable – than new stuff, so viewers must assiduously visit the Website to keep updated). Furthermore, given the viewing path on the video response page and of the ‘Play all’ feature (for both, from the most recent response to the oldest), viewers are likely to perceive as first-time used semiotic resources the most recent ones and, in turn, as ‘already done, nothing innovative’ the oldest videos, which, nonetheless, were posted earlier in time. Here the affordances of the interface configure in a different way the viewers’ perceived firstness (= the newest) in respect to the participants’ one (firstness = the oldest).

### **3 THE ‘MOST RESPONDED’ TOP CHART: AFFORDANCES AND PRACTICES**

On January 25 2007, under the title ‘Lotsa Cool New Stuff’, another blog entry by *The YouTube Editors* announces some new functionalities, and, among them, the introduction of the ‘Most Responded’ videos top chart (see Fig. 13):

VIEW VIDEOS THAT HAVE THE MOST RESPONSES: Under the Video

browse tab, there's a new page showing videos that have the most responses. Go to the Most Responded >> All Time to see a great set of videos that have inspired other users to post their own clips.

([http://www.youtube.com/blog?entry=9\\_wU0qhEPR8](http://www.youtube.com/blog?entry=9_wU0qhEPR8) Retrieved May, 26 2008)

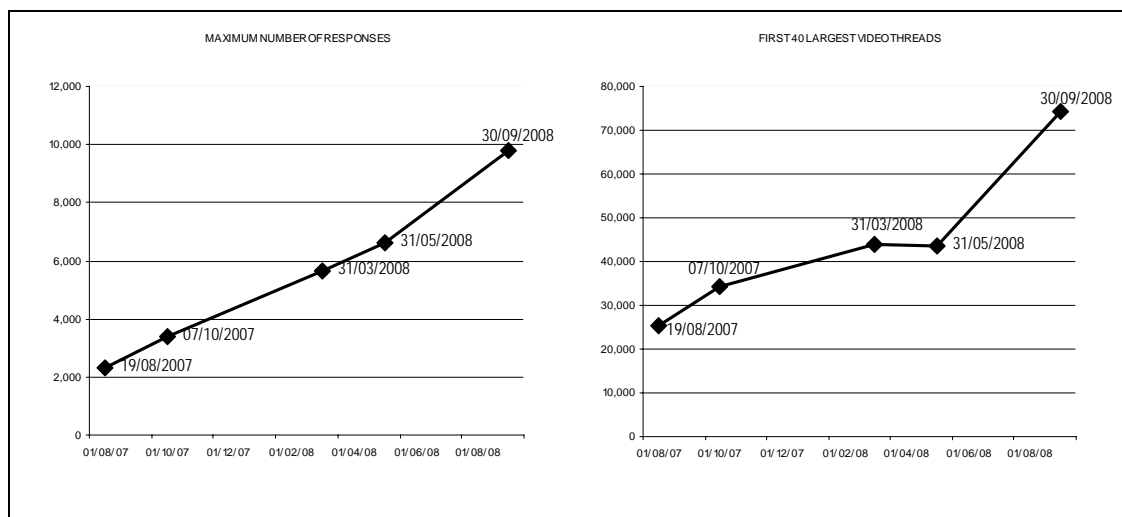
Here, probably motivated by the fact that these are the most responded videos, the use of the response option is defined as a consequence of 'inspiration' (rather than, as before, to create a 'dialogue' or 'reply'; cf. Sections 2.1 and 2.3). The label is an affectively connoted equivalent for 'prompt', the category used here to define the function of a video as a complex sign (and of one or more of its elements) of stimulating or provoking a video response (and of one or more of its elements).



Fig. 13 The Most Responded videos (All Time) top chart page.

### 3.1 Type of Most responded videos

As the label indicates, the 'Most responded videos of all time' are the videos that have received the largest number of responses. These are the videos which start the largest video-threads currently available, so that they initiate the largest existing instances of video-interaction. As mentioned in Chapter 3 (Section 2.1.1), the first two pages of the most responded videos of all time have been monitored over 14 months, since August 2007; each page displays 20 video thumbnails. The data were recorded on five different days, i.e. on 19 August 2007, on 7 October 2007, on 31 March 2008, on 31 May 2008 and on 30 September 2008. As Fig. 14 shows, the number of responses of the first video charted has quadrupled in 14 months (graph on the left of Fig. 14) and so has the total number of responses of the 40 videos which initiate the largest video-threads (graph on the right of Fig. 14).



**Fig. 14. Number of responses of the first most responded video (left graph) and of the first 40 Most Responded videos (right graph) over time.**

A clear categorization can be made of the type of videos which have collected the largest number of responses over the 14 months, i.e., which, according to the *YouTube* editors, have mostly ‘inspired’ others to post videos.

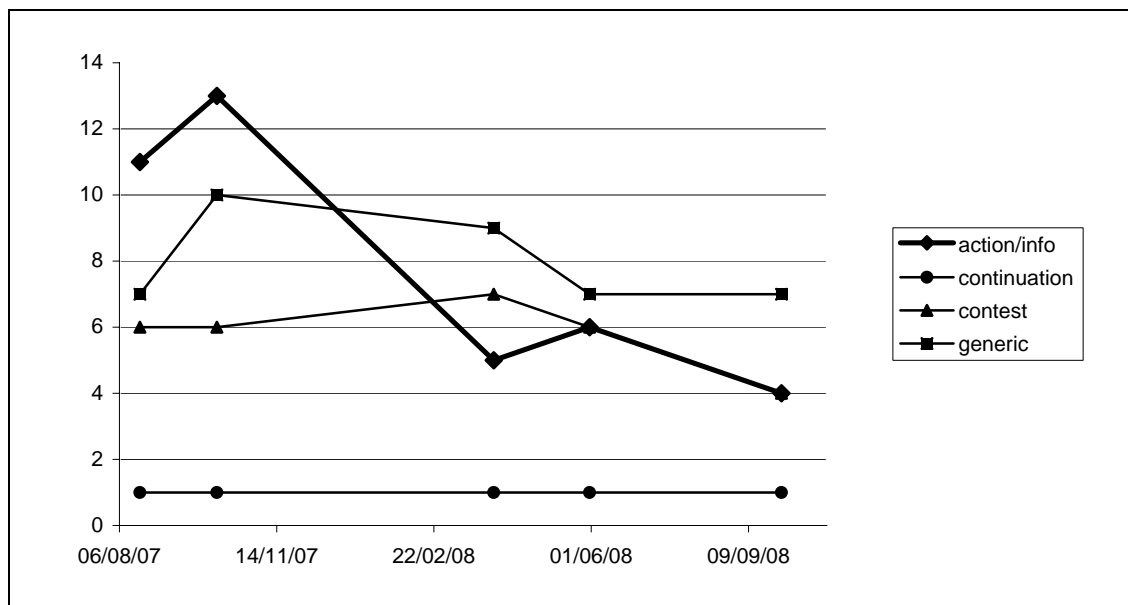
### 3.1.1 Video requests

In all five monitored days, about half of the videos displayed on the first two pages (i.e., the first 40 most responded videos of all time) are video requests, i.e., videos which ask explicitly for responses. More specifically, these video requests are 25 in August 2007 and 30 in October; they are 22 in March 2008, 20 in May and 16 in September. Although video requests are still the most frequent type of most responded videos, in over a year period their number decreases (from 63% to 40%).

Video requests can ask to provide information (e.g., the video titled ‘Where Do YouTube?’) or to perform an action (e.g., ‘Insult fellow (You)Tubers’), or else, to continue something that is represented in the initial video (e.g., ‘Continue this story’). They can also launch thematic contests or database (e.g., ‘YouTube Talent Contest’), they can ask for responses whose content is not topic-specific (e.g., ‘the best of youtube’, or ‘the worst of youtube’), or explicitly ask for as many responses as possible no matter the content, so to make the initial video the most responded one (e.g., ‘everything’, ‘respond to this video’). Some video requests also promise a follow-up in terms of a remix video (or mashup)<sup>60</sup> of a selection of shots of the

<sup>60</sup> ‘Remix vlog’ is a well established genre among videobloggers; indeed, ‘remix vlog – or mashup – is a video that takes other people’s clips and mixes them together, usually with an artistic (often musical) goal [...] The idea comes out of the DJ mix movement’ (Bryant, 2006: 319); cf., also, Dedman and Paul (2006) and Verdi et al. (2006). The novelty in video-interaction is that the mashup now resumes a communication thread of videos made *purposely* for the interaction.

responses (similar to the posting of a summary in forum discussions), or in terms of a video proclaiming the winner, for the ones which launch contests.



**Fig. 15. Distribution of the typology of video requests in the Most Responded top chart.**

As Fig. 15 shows, during the 14-month monitoring period, the number of requests for information/action have decreased, so have the contests, while generic video requests (i.e., requests asking to post responses irrespective of their content) have remained stable in frequency. In other terms, topic-specificity of video requests has decreased in the top chart in favour of ‘genericity’. Given that the number of responses of the charted videos has considerably increased along time (cf. Fig. 14 and its related discussion above), topic-specificity is harder and harder to make a video charted.

As anticipated in Chapter 2 (Section 4.5), the interaction occurs primarily between each response and the video it responds to. Although asynchronous and not in co-presence, in video-threads started by a video request, the exchange structure is similar to face-to-face group interaction where a leader asks the participants to introduce themselves (such as in seminar openings or in therapy group sessions) or sets a theme/topic to be developed by the participants (as in some dance, music or acting classes; or in free-style dance, jazz and rap sessions)<sup>61</sup>. In sequence, each participant gives one’s own contribution, modulating it according to the initiator’s request and avoiding repetitions with the previous ones, so as to differentiate oneself.

<sup>61</sup> Another themed participation is the one taking place in science-project and choir contests. Although also here each participant can watch the other contributions, differently from video-interaction, she has already designed one’s own and is likely not to have the time to revise or adjust it according to the previous ones. In jazz sessions, the initiator gives the theme but it is rather developed by each contributor who does not only address the initiator but also the next participant (similarly, in this respect, to a relay race).

No participant intervenes on other participants' contribution, they all reply only to the initial request, yet each contribution *indirectly* interacts with the other, in the sense that, even if not addressing one to another, all previous contributions have effects on the ones that follow. Indeed, as mentioned in Chapter 2 (Section 4.5), a secondary type of interaction does occur among responses, since each takes into account the previous ones, trying to avoid repetitions while keeping the same pattern.

In the video-threads analysed here, this 'side-interaction' among responses is evidenced by the responses' impressive representational variation, which is not at all accidental; cf., for example, a comment to a 'Where Do YouTube?' video response lamenting having used its same resources (i.e., *Google Earth* and *2001: A Space Odyssey* soundtrack):

Commenter:           oh my god! i just made a video for where do youtube from??  
                              and I used google earth and the same song!!! I wanna kill  
                              myself!!! Hahahah

Uploader's reply:    I just saw it, its cool. somehow I feel like this earth tis too  
                              small hehehe

The first video author somehow tries to comfort the commenter and answers the comment both giving a positive evaluation to the other's video and showing solidarity (cf. 'somehow I feel like this earth tis too small', i.e., we all share the same difficulty in finding distinctive representational means in videos, because we share the same limited means for producing them):

In sum, what responses have in common, or better, the regularity of patterns emerging from them, as well as what differentiates each of them – the extent of variation among them – are the formal effects of this interaction, i.e., the elements in the products (video responses) through which the process (interaction) can be detected. Eventually, the possible follow-up is the initiator's subjective selection, which is presented as an objective resume of the whole interaction.

Therefore, in the light of the analogy with group introductions, the interactional structure is similar to a 'circle' around the initial video, as represented in Fig. 16.



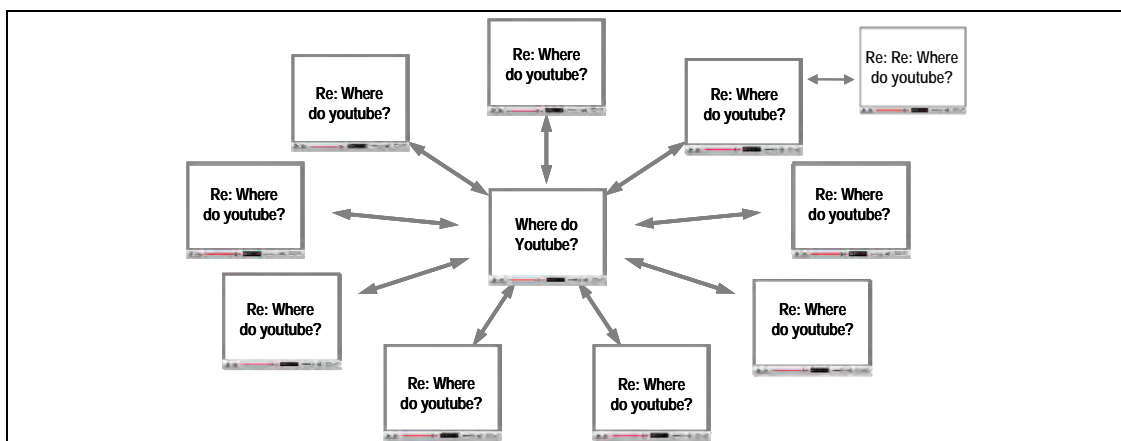


Fig. 16 The shape of the thread which is built ‘around’ the initial video.

In the analysis in Chapter 5, two threads are initiated by two of these most responded video requests; one is the thread that starts from the video titled ‘@----Where Do YouTube----@’; the other – used as side-reference for punctual comparison (cf. Chapter 3, Section 2.2) – is the thread that starts from the video titled ‘Why Do You Tube?’. They both have a specific question as the topic of their request; the former asks (You)Tubers to provide information on where they (You)Tube (from) and post it as video response, while the latter asks (You)Tubers to provide the reasons why they (You)Tube. As the related analysis shows, judging from the type of responses in the thread, both initiators exercise a weak control on the threads (which are constituted also by some totally unrelated responses). They both promise a follow-up, but only the first one has actually posted it. Judging from the type of selection of the shots that have been included in the video-summary, the first initiator seems to be quite open to differentiation, which is nevertheless presented as a vital element within a harmonic whole, while he himself selects and transforms the resources used in the responses, by recontextualizing them according to his interests. This generally acknowledged and welcomed misinterpretation of the interactional exchanges once again proves that successful communication disregards the interactant’s mutual understanding in the practice of video-interaction (cf. Chapter 5, Section 5).

### 3.1.2 Prompting videos

Among the Most Responded video, the ones which do not ask for any response can however be understood as ‘demanding’ a response, i.e., they prompt a response without explicitly requesting it. They are here labelled as ‘prompting videos’.

Judging from the type of prompting videos and the type of responses they collect<sup>62</sup>, three different types of prompt can be identified:

<sup>62</sup> Indeed, as discussed in Chapter 2 (Section 4.1.4), the range of prompts can be inferred only from the type of prompts that are actualized (taken up) in the responses.

- prompting of discussion/debate: their ideational content is controversial so that it raises thorny issues, like the video titled ‘Atheist Paradise’ where the issue of atheism<sup>63</sup> is raised (by means of the protagonist singing a text on the topic and performing mildly sacrilegious acts in religious places);
- prompting of emulation: their ideational content is a more or less extraordinary performance, so that it prompts imitation and alternative performances, i.e. they are understood as a challenge to imitate (and maybe surpass/outrun) them. This is the case of the video titled ‘guitar’, where a (You)Tuber plays Johan Pachelbel with his guitar, giving way to a series of imitations by reiteration/innovation; for an insightful analysis of this video and its follow up, cf. Burgess, according to whom the video has started a ‘cycle of imitation, adaptation and innovation’ (forthcoming);
- prompting of spoofing: their ideational content is so ‘bizarre’ that, according to (You)Tubing practice conventions<sup>64</sup>, they ‘demand’ a parody; this is the case of ‘Miss Teen USA 2007 - South Carolina answers a question’, which is the clip of an excerpt of the TV show where the Beauty contestant answers a question in a totally incoherent way (which proves again an old male discourse on the ‘lack of intelligence of beautiful *chicks*’), or the case of the video ‘Leave Britney alone’ in which a (You)Tuber weeps and cries, pleading for the cease of malicious gossip around the singer Britney Spears (cf. Chapter 6, Section 1).

For the sake of explanation, specific videos have been brought here as examples of the three prompts. However it must be noticed that the three prompts do not correspond to three separate types of videos, but, as the notion of prompt is here understood, these are rather different aspects of the meaning of videos which can potentially provoke differently modulated responses.

In the prompting of discussion/debate, what is taken up by the responses is the ‘ideational meaning’ represented in the videos (the thorny issue it raises). In the prompting of competition, what is taken up is the ‘interpersonal meaning’ presented in the videos (the participant’s will to prove they can perform better). In the prompting of spoofing, what is taken up is the ‘textual meaning’ of the initial video (indeed, a parody is recognizable as such because it is a variation on the basis of salient formal elements of the original). As happens to the three Hallidayan meta-functions (Halliday, 1978), also the three types of prompts can be (and usually are) combined in a video, so that a thorny issue can be represented by means of an outstanding amateur performance (as the ‘Atheist Paradise’, whose protagonist sings

---

<sup>63</sup> Atheism and religion are widely discussed topics on *YouTube*, as testified also by other videos ranking among the Most Responded ones: e.g., the video request ‘The Blasphemy Challenge’ (4<sup>th</sup> in August 2007 and 13<sup>th</sup> in September 2008), which has also given rise to ‘The Blasphemy Response Challenge’; the video request ‘A Massive Islam Campaign – 2008’ (10<sup>th</sup> in September 2008); and the debate-prompting video ‘Appeasing Islam’ (40<sup>th</sup> in September 2008).

<sup>64</sup> This is also ‘motivated by the specific ethics of [...] internet subculture, oriented around absurdist and sometimes cruel frathouse humour’ (Burgess, forthcoming).

his original rap song), and parodies and debating responses can be equally prompted by the same controversial video (indeed most threads are composed of both types of responses). Eventually, a parody can be an instrument of critique of a thorny issue and is also an alternative performance imitating the ‘original’ in a playful way. The latter case is testified by the responses to the highly popular video ‘Chocolate Rain’, which is an amateur performance of a song. The singer is a very young boy with a peculiar bass voice. The video has soon become viral<sup>65</sup> and has led the way to a number of alternative performances and parodies; cf. Burgess (forthcoming):

the *uses* of ‘Chocolate Rain’ as part of participatory culture ended up far exceeding the original intentions of either the original producer or the original disseminators. There was a relatively brief but highly creative flurry of parodies, mashups and remixes as Chocolate Rain’s popularity spiked. These derivative works reference ‘Chocolate Rain’ by imitating or re-using parts of it, and frequently combining them with many ideas from other sources, building on layers of knowledge built up in previous internet ‘phenomena’ as well as broadcast media fandom (like *Star Wars*). (Burgess, forthcoming)

The analysis in Chapter 6 focuses on a thread initiated by one of these prompting videos, which do not explicitly request to post responses. It is the thread started by the ‘Best video EVER!’, authored by *ChrisCrocker* a highly discussed *YouTube* celebrity, who is featured facing the camera, blinking twice and smiling to the viewer. The thread initiator does not exercise a great degree of control over the thread, which is composed by both variously related responses and by totally unrelated ones. As evidenced by the analysis results (Chapter 5, Section 6), judging from the composition of the thread, it seems that responses can be prompted by:

- (a) some salient elements represented in the initial video;
- (b) some salient elements related to the (You)Tuber (her persona or her (You)Tubing history, i.e., any of her previous videos);
- (c) some salient elements implied in the practice of (You)Tubing, i.e., one of its main purposes, which is, as epitomized by *YouTube*’s slogan, to ‘broadcast

---

<sup>65</sup> On viral videos Burgess observes:

As this example shows, there is much more going on in viral video than ‘information’ about a video being communicated throughout a population. Successful ‘viral’ videos have textual ‘hooks’ or key signifiers, which cannot be identified in advance (even, or especially, by their authors) but only after the fact, when they have been become prominent via being selected a number of times for repetition. After becoming recognisable via this process of repetition, these key signifiers are then available for ‘plugging into’ other forms, texts and intertexts—they become part of the available cultural repertoire of vernacular video. Because they produce new possibilities, even apparently pointless, nihilistic and playful forms of creativity are contributions to knowledge. This is true even if (as in the case of the ‘Chocolate Rain’ example) they work mostly to make a joke out of someone [...] the dynamics of viral video could be understood as involving the spread of replicable ideas (expressed in performances and practices), via the processes of vernacular creativity, among communities connected through social networks. (Burgess, forthcoming)

yourself”, or, in other words, to get one’s video viewed. In this latter case, posting one’s video as a response to a very popular one, allegedly enhances the chances to reach a larger audience<sup>66</sup>.

Any response can take up one or more of these three prompts. If the response is only prompted by (c), i.e., it gives no clue as to its relation to the first one or to its uploader’s persona, it functions like traditional ‘spam’, in that, like all forms of spam communication, its purpose goes beyond interacting with the semiotic act it responds to and is primarily directed towards reaching a wider audience. Nevertheless, unlike usual spam, here the link of response has been approved by the thread initiator, so that it cannot be considered an instance of disruptive (unsuccessful) communication. In other words, the different interests of the thread initiator (having as many responses as possible) and of the video-respondent (linking her video to a popular one so as to achieve visibility) are compatible, so that a semantically incoherent exchange is successfully realized in that it fulfils the participants’ interests.

### 3.1.3 Anomalous most responded videos

From time to time, among the Most Responded Videos of all time, some ‘anomalous’ instances are charted. These are videos which neither request responses, nor prompt any response in a direct way, whose threads are composed almost exclusively by totally unrelated videos. Their appearance among the Most Responded Videos of all time is not the product of any prompt-response relation between the video and its responses, but rather of some speculative operation.

One of these instances, which appeared in the Top Chart in August 2007, is the video ‘me getting used to my webcam’, uploaded by *AngelaSWilliams*. The video belongs to the videoblogging genre, i.e., with the typical modality of a Webcam recorded video: low definition, bad sound quality, uncertain lightning, the (You)Tuber filmed indoor, facing the camera at a mid-close shot and speaking in an apparently unplanned manner – i.e., ‘ranting’ – with many syntactic changes and random content etc. The ideational meaning of the video is nothing but ordinary, like thousands of other videoblogs, or even worse, since the bad sound quality prevents most of the ‘ranting’ from being intelligible. Hence its appearance on the most responded top chart was quite inexplicable.

As some users revealed later in their videos (e.g., the video titled ‘AngelaSWilliams the Cheater’s Secret revealed Part 1’), its massive number of video responses was the result of a huge sending of text messages posted from *AngelaSWilliams* account to random *YouTube* profiles, inviting them to subscribe to her channel and to post responses and comments. As a consequence of the policing activity by (You)Tubers, the user was eventually blocked by *YouTube*, the video removed and the account

---

<sup>66</sup> However, in this regard, cf. Benevenuto et al. (2008a), whose quantitative study evidences that there is no relation between number of views of a video and its being posted as a response to another one.

terminated.

Policing practices among (You)Tubers (enabled and fostered by the interface), such as ‘flagging’ the video as ‘inappropriate’, ‘blocking the user’, or even reporting a violation of the terms of use to the Website owners, usually detect and stigmatize these ‘speculative operations’, which are sometimes publicly exposed by means of videos which unveil the ‘scam’, as in this case. This policing activity and public accusation generally results in the so-proclaimed ‘disruptive’ users being blocked or their accounts terminated (i.e. the ban from the community) and their videos being removed.

The same policing activity can be further taken up by the Website owners and can result in changes in the affordances of the structure (i.e., changes in the interface options and functionalities). Indeed, after the ‘AngelaSWilliams’ case, in order to prevent the practice of spamming messages, the interface has been enhanced so that, after sending a certain number of comments, users are asked to solve a CAPTCHA (a computer-generated image of a text which artificial intelligence cannot read) prior to be able to send any further comment.

Nevertheless, prompted by the great stress on ‘visibility’ with which the activity of (You)Tubing is endowed by the Website in particular (i.e., its slogan ‘Broadcast Yourself’, its honours and top charts system, etc.), and by the ‘Society of the Spectacle’ (Debord, 1967) at large, these practices keep going and renewing themselves, by creating new user accounts or by using innovative strategies when the previous ones are prevented by structural changes.

Thus, another type of anomalous ‘hits’ in the Most Responded top chart are videos which start threads composed by a peculiar type of video responses posted by a very limited number of *YouTube* profiles. These are tens, or even hundreds, of very short (1” or 2” long) video responses, all with the same title, sometimes also featuring the same unique shot (or, anyway, with the same multimodal deployment, in terms of colour palette and elements displayed), so that, even from their thumbnails in the video responses’ page, it is immediately evident that these responses are minimal variations generated out of the same matrix. This is a new practice that ‘spammers’ (or, better, ‘flooders’; cf. footnote 28) have devised in order to overcome the constraint that a video cannot respond to more than one video. In the data, this practice was first monitored in August 2007, in the thread started by the video ‘INSULT fellow YOUTUBER’, which, by that time and thanks also (but not exclusively) to these very short repetitive videos, had reached the 6<sup>th</sup> position in the Most Responded top chart, with 1,176 responses. This practice is quite controversial among (You)Tubers, so that, on the one hand, (You)Tubers’ policing activity has led both the thread initiator’s profile and the responding one to be suspended; however, on the other hand, along the monitoring period, videos in the top chart which are responded mostly by few uploaders posting a huge number of short and almost identical videos have augmented. Indeed, while in August 2007 only one of the first

40 most responded videos had this type of responses, in October they were two, five in March, six in May and, significantly, 16 in September 2008.

Thus it seems that the Most Responded Top Chart is more and more characterized by a very distinctive (sub-)genre of video-interaction, in which interaction by means of videos is used as a means of gaining visibility, rather than for interacting on topic. This ‘flooding’ interacting practice is getting acknowledged by (You)Tubers, so that in September 2008 one of the most responded videos<sup>67</sup> is a request launching a contest which promises a reward in terms of number of subscribers to the (You)Tuber who will post the greatest number of responses (no matter the content).

From the above, it becomes evident that the relation between affordances, uses and policing practices is rather complex. On the one hand the exploitation of the affordances according to varied (You)Tubers’ interests gives rise to unexpected practices, like the ones named here as ‘anomalous’; on the other hand, these practices are sanctioned by some (by exploiting the policing affordances on the interface), while fostered by others. Time will tell how these contrasting interests will influence the Website editors in changing the affordances of the interface.

### **3.1.4 Related flooding responses**

As usually happens, when affordances are used for different purposes by different interactants there is never a straightforward directionality between a functionality and its effective uses, and interactants often create innovative ways to use the available resources according to their aims. So, the above discussed practice of flooding a video with numerous very short videos is used by some (You)Tubers to make meaning in a peculiar way. These (You)Tubers exploit the interface functionality of the thumbnails displayed in the responses page, so that meaning is made by the sequence of the thumbnails posted by the same user in succession rather than from the content of each video (which is nothing more than a few-second still, and thus it is the video equivalent of the image displayed in the thumbnail).

This phenomenon occurs for example in a series of videos posted by the same profile as responses to the video ‘Èric and the Army of the Phoenix (1/5)’<sup>68</sup>, which was charted as the 10<sup>th</sup> most responded video in the Top Chart in October 2007 with 1,280 responses. Since 31 March 2008 it has reached the first position in the top chart (on the last recorded day it had collected the impressive number of 8,340

---

<sup>67</sup> This is the video ‘Contest: Win LOADS of Subscribers!’, uploaded on July 11 2008, which on 30 September 2008 has collected 608 responses posted by no more than 30 usernames.

<sup>68</sup> As anticipated in Chapter 3 (Section 2.2), this thread is analysed here, instead of dedicating a separate chapter of analysis to it, since it is not precisely the focus of the analysis of the texts of video-interaction. Indeed, as discussed hereafter, it does not make meaning by virtue of the contents of the videos, but rather by virtue of their thumbnails displayed on the page of the responses. Nonetheless it is to be considered as part of the texts of the analysis (together with the above discussed ‘AngelaSWilliams’ thread).

responses). As introduced in its description, the initial video is the first of a five-part documentary account of

An incredible but true story: Spanish authorities prosecute child for terrorism when he e-mails companies requesting labelling in Catalan language, using Phoenix monicker from Harry Potter books. Police accuse him of organizing an Al Qaeda cell. Case goes all the way to Spanish High Court.

When looking at the related pages, it is immediately evident that most of the responses are posted by a very limited number of uploaders. Judging from their thumbnails and titles, responses seem to be related to the initial video by a semantic field which could be termed as ‘Catalonia’; so, some videos are titled ‘literatura catalana’ and have thumbnails portraying writers or book covers, others are titled ‘paisos catalans’ and have thumbnail images of places, yet others are titled ‘joan miro’ and have his paintings as thumbnail images etc.

The thumbnails of the responses from no. 1,531 to 1,631, displayed on pages 85 and 84, are all images of flags, i.e., recognizable by means of the conventional shapes and colours for flags (Fig. 17 shows the first thumbnails displayed on page 85 as retrieved on 1 June 2008).

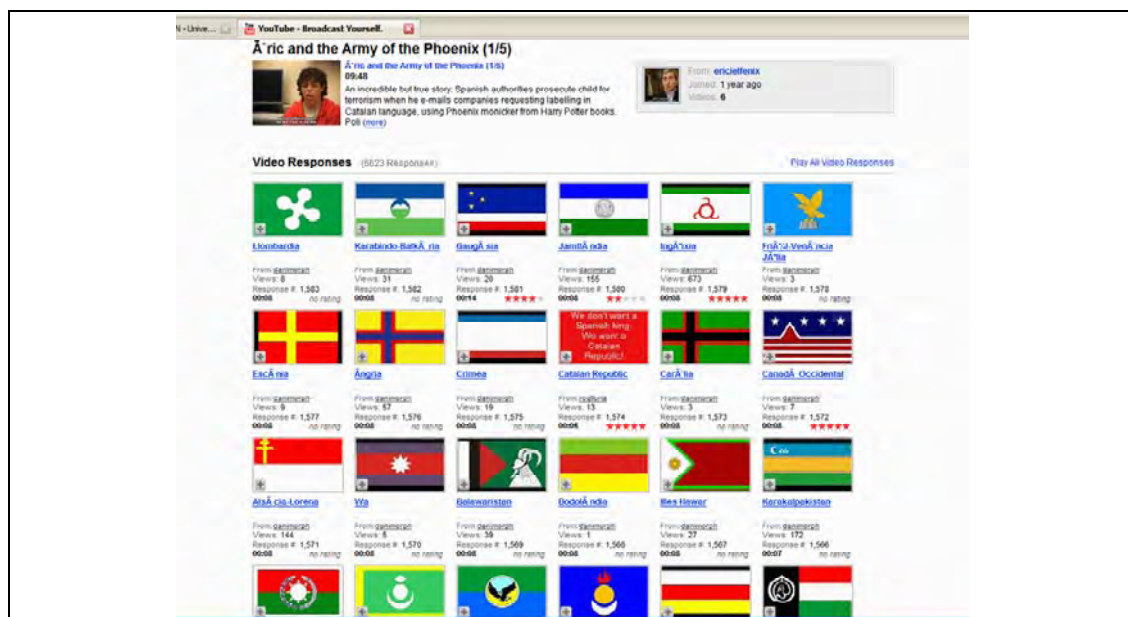


Fig. 17. An example of flooding of related-responses.

The title of each video is the (Spanish) name of a region which has some open issues concerning independency claims (e.g., ‘Lombardia’, ‘Darfur’, ‘Mongolia interior’, etc.). Some names are not widely known at all, at least to European viewers (e.g., ‘Balawaristan’ or ‘Wa’), but the viewer can infer that they all refer to these would-be

independent states by semantic association with the ones she knows<sup>69</sup>. This associates further with the ‘Catalan’ (secondary) topic of the initial video, and to that of the ‘Catalunia’-related responses, thus the meaning of that sequence of responses results in something like: here are all the countries, which, like Catalonia, claim for independence but have not achieved it yet. This meaning is made by watching the thumbnails on the responses page, not by watching each video response. All these videos are very short; their duration, which is also displayed on the thumbnails page, spans from 2” up to 9”; most of them last 8”. By clicking on the thumbnails, one can access to the response video page; the video content is structured in the same way in all of them: they are composed by two different shots; first is a blue screen, on which, typed overlaid in white fonts is the name of the country in Spanish (corresponding to the video title) and, in brackets, the name in the autochthon language – e.g., ‘Alsàcia-Lorena (Elsaß-Lothringen)’; second appears the still image of the flag which lasts for the remaining duration of the video; the video description reads as: ‘Bandera de +[Name of the country]’ and the tags are ‘[Name of the country]’, ‘independence’, ‘flag’. Of course, one can watch and make meaning out of each video separately, but a further, more ‘relevant’ meaning is given by watching the sequence of the thumbnails displayed in the responses page of a video of a Catalan boy charged with terrorism by the Spanish authorities.

Other examples which exploit the same device to make meaning include a series of thumbnails with a few words typed overlaid onto a coloured screen, or a series of thumbnails reproducing the same drawn image with just slight differences in each of them, so that the sequence of the thumbnails functions like the storyboard or the different shots of a cartoon animation. Given that the responses’ thumbnails are displayed in reverse chronological order of posting, according to the Western reading conventions the reading path on the page is from left to right and from top to bottom. So, if the meaning is not a list of entities (as in the case of the flags) but rather a sequence of events (as in the case of the written text thumbnails or in the sequence of drawings), the uploader ought to reverse the order of posting for it to be displayed in a sequence that follows the conventional reading path. Alternatively, only experienced *YouTube* viewers, who know that the first-posted response is displayed at the right bottom of a page, can reverse their conventional reading path according to the reversed chronological order displayed by the interface.

The flooding of related responses is an innovative semiotic practice which makes meaning through still images by means of videos on a video-sharing Website. In terms of interface affordances, it exploits the spatial-sequencing layout of the responses’ page rather than the streaming of videos. In terms of meaning production, it reverses (or better, unwinds) the cartoon/filmic process; indeed the latter makes

---

<sup>69</sup> For example, given my Italian background, I can recognize ‘Lombardia’ and ‘Friuli Venezia Giulia’ as regions of Italy; I associate these video titles with the type of image in the thumbnail, which, by means of shapes and colours, is unmistakably conventional for a flag; ‘flag’ means ‘nation’; I can further recognize some other video titles as names of regions whose independent claims are often mentioned in the news, like Tibet and Kurdistan.



meaning through the dynamicity created by a fast succession of a sequence of still images, whereas the meaning of this flooding of related responses is given by the spatial sequential display of still images, which are the result (thumbnails) of dynamic texts (videos).

Eventually, in terms of social affordances, the flooding exploits the high visibility of the Most Responded top charted videos in a quite complex way. Indeed, the massive flooding of responses makes the initial video charted among the most responded ones. On the top chart, it is very likely that viewers go to the thumbnail pages to see what type of responses have been posted to these charted videos, while it is very unlikely for them to watch each of the several thousand responses. Therefore, while contributing to make the initial video appear on the top chart, the flooding simultaneously uses a social value (i.e., the interest that viewers have in top charted videos) and practice (viewing thumbnails of the most responded videos, instead of playing the whole very long thread) to make their own meaning and distribute it.

In sum, the flooding of related responses is an innovative semiotic practice on *YouTube* which, by exploiting the affordances of the medium, achieves (at least) two aims at once, i.e., (a) it contributes to the popularity of the initial video, by enlarging considerably its number of responses, and (b) it makes its own meaning, by exploiting the page which displays the responses thumbnails. Furthermore, given that the flooding is topic-related to the initial video, its texts cannot be considered as 'spam' and so they are hardly subject to policing activity by (You)Tubers.

## 4 CONCLUSIONS

This chapter has focused on video-interaction as process. In the first section of this chapter, the structural characteristics of video-interaction have been introduced and discussed. These are: (embodied and disembodied) multimodality, homogeneity and bidirectionality, publicity, asynchronicity, disembodiment, online, distance, multiple mediation, and corporate interface distribution. This combination of features makes video-interaction a distinctive form of communication and their analysis, by means of its similarities and differences with other communicative processes, has led to map the place of video-interaction in the contemporary semiotic landscape.

Sections 2 and 3 have discussed the introduction and affordances of the video response option and of the Most Responded videos top chart. They have further focused on the practices that have developed through the use of the interface affordances according to the participants' diversified interests, giving rise to unexpected uses of the functionality which can lead to changes in the affordances of the structure itself. This cyclic process brings a step further the notion of 'appropriation' of Cook et al. (2008) for the use of mobile technologies, which is defined by the innovative uses of the device according to the user's purposes:

We define appropriation as exploration, accommodation, assimilation and change for and in context-governed meaning-making with users/learners negotiating and evolving practices and meanings in their interaction with other users/learners, technologies and information. (Cook et al., 2008)

The video response option was introduced on the interface in response to an unexpected use that participants were making of the related videos functionality. While the latter is based on automatic linking according to intertextual relations of user-generated tagging, the former works on the participants' intentional interaction.

Censorship powers given to the participants in filtering out incoming attempts at interacting make only successful instances of interaction observable on the Website, so that only the products of the interaction are public, from which the process must be inferred. On the basis of their textual organization in relation to the video they respond to, it is possible to infer how a response was created (whether *ad hoc* for the interactional exchange or in a previous time). A further interface enhancement makes all the responses to a given video an entity to viewers, so that the interactional thread is now likely to be viewed also by third parties, other than the participants in the interaction. Since the most recent responses are (dis)played first, it is more likely that 'firstness' and 'newness' are associated meanings when watching the thread.

On the basis of a monitoring period of 14 months, the analysis of the most responded videos has attempted a categorization of the type of videos which initiate the largest threads of video-interaction. These can be:

- (a) more or less topic-specific video requests (which solicit responses);
- (b) prompting videos (which, without asking for them, prompt responses by virtue of their ideational, interpersonal and/or textual meaning);
- (c) videos which reach the top chart because of speculative operations;
- (d) videos which are flooded by short (related) responses posted by the same uploader.

With the impressive increase of the number of responses of the top charted videos, topic specificity has given way to genericity, while video requests and prompting videos have given way to flooded ones. At present it seems that the flooding practice (especially if it is made of videos which are topic-related to the initial one, so that it cannot be policed as 'spam') is effective in making a video charted and in making meaning by virtue of the thumbnail arrangement on the responses page.

The analysis of the Most Responded top charted videos and their threads must not lead to generalizations onto the whole phenomenon of video-interaction or on how the functionality is used at large. It is simply an indication of how participants' create and transform their practices in response to the affordances of the structure (i.e., limits and possibilities offered by the resources) and to fulfil diversified purposes, many of which are socially fostered, i.e., prompted by powerful sources, like the Website owners and the way they shape the interface in terms of promoting

popularity and visibility indexes.

After discussing and analysing video-interaction as process, it is now possible to analyse the texts of the corpus and the ways in which they relate one to another, i.e., how they realize the participants' interest-driven prompt-response relation. While exemplary threads which start with videos typified in (c) and (d) above have been briefly analysed here, the threads initiated by videos of the types (a) and (b) are devoted a separate analysis in each of the two following chapters.



## **CHAPTER 5**

### **ANALYSIS 2/3:**

### **‘WHERE DO YOUTUBE?’ THREAD**

*‘it is the speech-act which, through its continuous circulation,  
propagation and autonomous evolution, will create the interaction  
between individuals or groups who are far away, dispersed,  
indifferent to each other’*

*‘interactions between supposedly dispersed and independent people  
who pass through the scene by chance’*

G. Deleuze, *Cinéma 2* (1985)

After examining the process of video-interaction in the last chapter, the present one opens the analysis to the texts of video-interaction, by describing and discussing a first video-thread which starts from a topic-specific video request selected among the ‘Most Responded Videos of All Time’. The thread is composed by its responses, by a video summary posted by the thread initiator and the responses to the video summary.

The thread is started by the video ‘@---- Where Do YouTube?----@’, which is a (topic-specific) request for responses. It is here examined, firstly by describing the composition of the thread (Section 1), then by analysing the initial video (Section 2), and by examining its responses according to the varied relatedness which they establish (by means of cohesive ties) with the initial video (Section 3). These will be analysed with reference to:

- a. how they relate to the topic-specific request by variously representing the elements of the answer to the topic-question (Section 3.1);
- b. how they construct formal relatedness with the initial video (Section 3.2);
- c. how the varied textual organization of the responses determines different degrees of relatedness in the thread (Section 3.3).

Relatedness is also the focus of the analysis of the sub-responses in the thread (Section 4) and of the responses to the video-summary (Section 6). Section 5 examines the video summary, with reference to how it selects and transforms the resources of the responses into a cohesive whole, thus presenting itself as a resume of the thread. Finally, Section 7 recapitulates the results and draws some conclusions.

# 1 THE THREAD COMPOSITION

The video titled '@---- Where Do YouTube? ----@' was uploaded on 2 March 2007 by *ChangeDaChannel*, a quite popular vlogger on *YouTube*; indeed his channel accounts for 6,248 subscribers and 265,470 views (data retrieved in February 2009). The video was very soon featured (i.e., selected by the Website editors and published on the homepage) and received 550 video responses in one month. In December 2007 it was charted as the 15<sup>th</sup> most responded video of 'all time'.

On 3 April 2007, a month after the posting of the initial video, *ChangeDaChannel* posted a video-summary, titled 'We tube (the World)', made of a selection of shots of the responses. The summary was uploaded with the following description:

Second Part to "Where Do YouTube From"  
With over 550 video responses from around the world and 3000 comments, this was some work...but fun. This shows the diversity on the one thing that brings us here, YouTube. Sorry I couldnt get everyone in. There was so many! Thanks alot for helping with this project. I hope to do some more stuff like this in the future:)  
Tracks by: Titus & Hesthesmartone  
Editing by: Changedachannel  
Video by: YOU:)  
Special thanks to everyone who took part & also Oaklynrecs, Gimmeabreakman, Pigslop

The video summary did not conclude the thread though and video responses kept being posted. In August 2007 the video was charted as 11<sup>th</sup> most responded one. In about 5 months since its upload, the initial video has collected 792 responses; 20 responses have received 33 responses in their turn, one of which has got a further response. Eight videos have responded to the video-summary and one of them has been further responded (Fig. 18 represents the thread composition).

Since then, the initial video description has been promising a second follow-up; meanwhile, the thread considered here is composed of 837 videos (an initial video, 835 responses and sub-responses, and a video-summary).

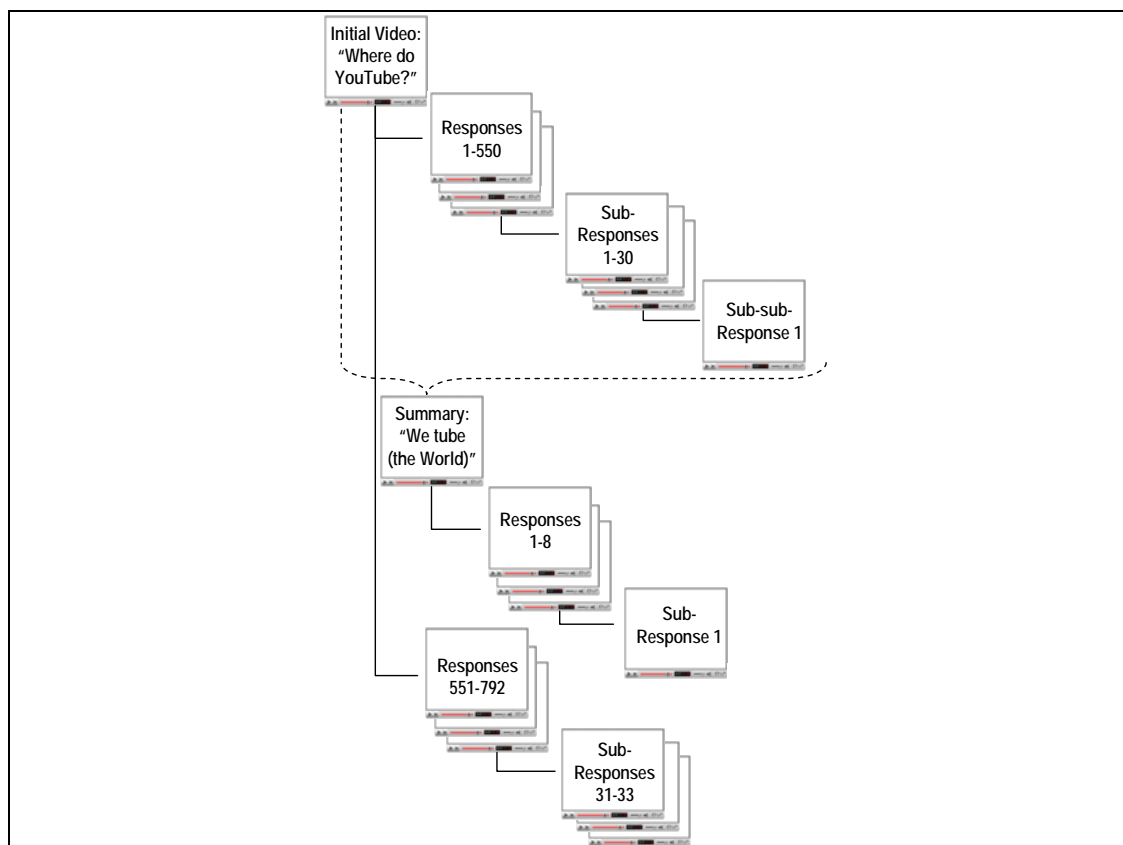


Fig. 18. 'Where do YouTube?' video-thread.

## 2 THE INITIAL VIDEO

1'.05'' long and shot in black and white, the initial video features *ChangeDaChannel* – his mouth covered – gazing at the camera. Several computers are in the background and a heavy-metal soundtrack accompanies the images. Fig. 19 shows the salient snapshots:

- a. *ChangeDaChannel* widens his eyes, points his finger to the typed-writing overlaid at the bottom of the screen, shakes his head and shrugs:

a subscriber said i talk too much in my videos...<sup>70</sup>

- b. while he shows pen and notebook, a second typing appears:

so I decided to do this video without talking ;P

<sup>70</sup> Since handwriting, colours and other graphic features (e.g., the display over the page) of the written material cannot be reproduced here, font types and sizes will not be reproduced either. Only lower and upper cases are reproduced as they appear in the videos.

- c. he starts to write;
- d. he gazes back at the camera and shows his writing:

I TUBE FROM CALI USA'

- e. he writes again and lifts his index finger (so as to ask viewers to wait);
- f. he keeps writing;
- g. he then shows this second writing:

WHERE DO YOU TUBE FROM?

- h. the screen splits and multiplies the shots;
- i. then the screen turns black with a typed invitation to post responses together with the soundtrack credits:

do a video response and let me know where YOU TUBE From  
'Vowels' Track by: Tituts Kilawattz

- j. a further typing appears with the promise of a follow-up and the invitation to join his channel:

And Ill do a sequence Thanks! SUBSCRIBE to catch the final "WeTube From"

- k. several seconds of soundtrack playing on a black screen<sup>71</sup> precede a colloquial closing in small fonts, composed of a question:

You still here?

- l. and an imperative which sends the viewer away (the illocutionary force being mitigated by the emoticon ';P' = 'I'm joking'):

Get CrackeN;P

---

<sup>71</sup> This type of closing and the 'wait' gesture with the index finger are quite recurrent in the corpus and they both function as suspense devices, so as to prevent the audience from abandoning the view of the video and move on to another one.





Fig. 19. Salient snapshots of '@---- Where do YouTube? ----@' video.

Specific signifiers in the video construct *ChangeDaChannel*'s identity and relationship with his viewers:

- *ChangeDaChannel*'s distinctive look (dark 'Slayer' t-shirt, bandana and cap, metal necklaces), which characterizes his persona in most of his videos, combine with the heavy-metal soundtrack in communicating 'toughness';
- the computers in the background present him as a practiced member of the cyber-community;

- his experienced<sup>72</sup> gazing at the camera ‘demands’ (Kress and van Leeuwen, 1996, 2006) a relationship with the viewer, who is called upon as addressee;
- the mid-close shot establishes an introductory distance with the viewer (close enough to make acquaintance, yet not too close to be invasive);
- the motivation of his mode selection (‘a subscriber said i talk too much in my videos...’) presents *ChangeDaChannel* as a popular videoblogger (i.e., his channel has subscribers), who pays attention to his viewers’ feedback and modifies his practices accordingly (‘so I decided to do this without talking ;P’);
- his anticipated follow-up promises a rewarding conclusion to respondents and functions as an indirect speech act (Searle, 1969) of invitation to ‘be part of something’;
- the closing suspense device establishes *ChangeDaChannel*’s confidence and playful authoritative role over the viewer, who has been watching the several seconds of black screen expecting something to happen, just to be eventually sent away.

At the interpersonal level of meaning, the signifiers in the video present a tough, experienced, popular and egalitarian (You)Tuber, who has a plan, listens to his viewers and learns from them, who authoritatively summons them, wittily plays with them and rewards those who participate in his project, presented as a way of making acquaintance.

*ChangeDaChannel*’s writing his response (‘I TUBE FROM CALI, USA’) before his request (‘WHERE DO YOU TUBE FROM?’) plays a complex set of functions. Firstly, it combines with the mid-close shot in specifying the situation of communication as a mutual introduction rather than an interrogation. Indeed, in introductions, we usually give information beforehand (e.g., ‘My name is...’), thus mitigating the potential threat of the question (‘What’s your name?’), which otherwise would sound as an interrogation. Since *ChangeDaChannel*’s request is not presented as an attempt at *grabbing* information, but rather at *sharing* some, there is consequently no threat in his question nor danger in answering it. Therefore, providing his answer beforehand mitigates the potential danger of disclosing information online, which is widely acknowledged by online participants in general and by (You)Tubers in particular, as mentioned also in some responses in the thread; cf., the following sung excerpt:

he asks just where do we tube / just think of me as out of town / just think of me as  
out of town / cause I don't want everyone to necessarily know / I've seen some  
unstable people here man

---

<sup>72</sup> See Adami (2008a) for gazing at the camera as signifier of videoblogging experience. Indeed, gaze trajectories other than to the camera are resources more readily available due to the medium affordances (essentially because, while video-recording, the recorded image is displayed on the pc screen, so that it is difficult not to look at one’s own image while it is being recorded).

Secondly, by enacting the production of the representation (i.e., his action of writing), *ChangeDaChannel* presents the response to his request as an easy task; anyone can write on a paper in front of the Webcam and, since *ChangeDaChannel* is doing it ‘live’, one can see how little time it takes. In sum, answering is both *safe* and *easy*; this strengthens the appeal of the invitation by mitigating the possible drawbacks (i.e., danger and effort) of the desired perlocutionary effect (Austin, 1962), thus functioning so as to encourage responses. Furthermore, in writing his answer, *ChangeDaChannel* sets handwriting as an exemplary mode, thus prompting a form-specific answer, while the content of his answer (‘CALI, USA’) sets the geographical location as exemplary information, thus prompting a content-specific answer.

In sum, *ChangeDaChannel*’s anticipated answer plays the interpersonal function of encouraging the action of responding, both in relation to its content (disclosing information is not dangerous) and its form (responding is easy), coupled by the ideational function of prompting the answer’s content (a geographical location) and form (the handwritten mode).

In other words, *ChangeDaChannel* designs his video by using the resources (available to him) which are the most apt to represent his meaning, according to his interest, so as to solicit as many responses as possible, in view of a summary representing where (in the world) people (You)Tube from. Arguably, the ultimate aim to which his representation ‘responds’ is to gain visibility on the Website as the editor of a worldwide collective video-project. Although no signifier explicitly signals this, achieving visibility is nonetheless always implied in any intentional semiotic act in all forms of public interaction (i.e., any public semiotic act is also the signifier of something like ‘I want you to acknowledge my presence’).

The next section discusses how responses take up differently these encouraged and prompted contents and forms and how, in so doing, they modulate differently their relatedness with the initial video and with its topic question.

### **3 THE VIDEO RESPONSES**

As listed in Fig. 20, out of 792 responses, eight videos are duplicates (i.e., posted more than once), two video clips are damaged (so that they cannot play) and one had been removed by its uploader before the data were analysed. 12 responses are set to private (only invited ‘friends’ can watch them); this way their uploaders achieve two simultaneous effects, i.e., their video (and the represented information) not to be publicly available (cf. Chapter 3, Section 5.1, and Chapter 4, Section 1.3), and *ChangeDaChannel*’s ‘friendship’ (in the case it were not already established) for him to watch their response.

<b>Video responses</b>	<b>792</b>
Duplicates	8
Damaged	2
Removed by the user	1
Set to private	12
<b>Accessible responses</b>	<b>769</b>

**Fig. 20. Accessibility of the responses.**

Out of the 769 accessible responses, 746 videos respond to the initial video by establishing various degrees of relatedness to it. Not only most of them answer the topic question, by representing – by various means – where their authors declare to (You)Tube from (thus establishing topic-relatedness with the initial video), but also (a few) others relate to the initial video only by means of formal elements, thus establishing relatedness with the initial video without however relating to its topic-request. Topic-relatedness is analysed in the following Section (3.1), before examining the role played by the mode in establishing formal relatedness in the thread (3.2). Eventually, Section 3.3 examines a continuum of relatedness on which the responses can be arranged according to the textual organization of both the semantic elements (the topic-answer) and the formal elements represented in the text and in the paratext.

Throughout the analysis, the notion of cohesion and cohesive ties (Halliday and Hasan, 1976) are used as a descriptive tool (yet adapted from verbal language to multimodal communication)<sup>73</sup>. Indeed, since the initial video has a topic-specific request, cohesive ties are the ‘litmus paper’ which helps distinguishing (various extents of) topic-related responses from others. In absence of these cohesive ties, it may be said that the exchange is (locally) incoherent, yet, as evidenced in the analysis to follow, this does not prevent it to be successful. As argued in Chapter 2 (Section 2.1), rather than dismissing them altogether, the notions of cohesion and coherence can be used to discriminate between coherent and incoherent exchanges. However, other higher-level notions – i.e., the prompt-response relation which may be instantiated by, e.g., attuning devices (cf. 3.2) or may result in marked textual organizations (cf. 3.3.3) – need to be used to account for all instances of successful exchanges in video-interaction.

The thread considered here is much more ‘cohesive’ and ‘coherent’ than the thread which is analysed in the next chapter. This is because the editor (i.e., the thread

<sup>73</sup> For an interesting attempt at adapting cohesion to dynamic texts, cf. Tseng (2008), who considers intra-textual cohesion (within a single text, i.e., a film), while in video-interaction cohesive ties are established – by the meaning-maker – between texts in an interactional exchange which involves (at least) two interactants. For a different approach to cohesion, cf. van Leeuwen (2005: 179-267).

initiator) exercises a greater extent of control on it. The editor has a clear project in mind (realized in the video-summary, which is maximally ‘coherent’ and which disregards the incoherent exchanges in the thread; cf. 5) and he monitors the thread accordingly (evidenced by the very limited number of video responses which are totally unrelated to the initial video; cf. 3.3.5). So most exchanges in the thread are broadly coherent in a traditional sense (i.e., most responses answer the question in a topic-related way), even if, in many notable cases, they are only locally coherent and not globally nor pragmatically coherent (i.e., they formally answer the topic question but do not provide the geographical information prompted by the initial video). Furthermore, this general cohesive – locally coherent – trend does not impede notable ‘exceptions’ to construct successful exchanges (i.e., topic-unrelated responses which construct successful exchanges just because they attune in form, cf. 3.2; or marked organizations of the response which would traditionally be considered unacceptable as an answer to the topic question, cf. 3.3.3). As often happens, limited in number as they may be, these exceptions are highly significant, in that they testify to the fact that video-interaction works on a notion of ‘successful communication’ which is rather different from the one contemplated in traditional communication theories. These exceptions become the ‘rule’ in the thread analysed in the next chapter, so that the tools of cohesion are even less usable in its analysis.

### **3.1 Topic relatedness: The answer to the question**

As said, 746 responses answer (formally and/or substantially) the question ‘where do you tube?’. They do it in different ways, so that various degrees of topic relatedness between the responses and the initial video can be established.

The elements which constitute the topic question, ‘where do you tube?’, are:

1. a locative interrogation marker (‘where’);
2. an agent referred to the addressee (‘you’);
3. an action (‘to (You)Tube’).

These are answered by an indication of:

1. a location;
2. [the addressed agent, i.e., the (You)Tuber];
3. [an action, i.e., (You)Tubing].

Since the locative circumstance is the focus of the question, the representation of a location is the essential item without which a response can be said not to answer the question (whether this is essential for the response to be perceived by viewers as related is another matter, cf. the discussion hereafter). Indeed, in presence of a represented location, the other two items (agent and action) are cohesively tied to the initial video by ellipsis, indicated in square brackets in the following verbal example:

A: Where do YouTube?  
B: [I (You)Tube from] Venice.

Even more, at a pragmatic level, the action of (You)Tubing is always implied in the uploading of a video, so that, given the existence of a video response, the action of (You)Tubing is factual. In other terms a video-as-a-text is a ‘performative’ (Austin, 1962) of the action of (You)Tubing. The action of (You)Tubing implies an agent, the (You)Tuber, who, if not stated otherwise, is referred to the username of the uploader (which is a piece of information available in the paratext of the video).

Therefore the representation of a location is the basic element for topic-relatedness (i.e., answering the topic question), while the (You)Tuber and the action of (You)Tubing can be cohesively tied by ellipsis, a semantic tie which is strengthened by the performative value of the video for both the action and the agent. Let us then analyse how the main element for topic relatedness is represented in the thread: the location. Let us first examine what responses answer (i.e., the content) before examining what they use (i.e., the form) to answer the question.

### 3.1.1 The geographical location

As anticipated, 746 videos answer the question ‘where do you tube?’ in different ways. Consistently with *ChangeDaChannel*’s prompted content, 708 responses represent a geographical location (Fig. 21).

Overall, 59 countries from all continents<sup>74</sup> are represented in the responses, although English-speaking and Western countries occur more frequently, USA above all. To these responses, two must be added which represent an unidentifiable geographical location (by means of a fast-forward zoom-in on *Google Earth* and of an unreadable handwritten paper; cf. the discussion hereafter). Although, understandably, no direct link can be established between the represented locations and their author’s physical ones, the distribution mirrors the ‘U.S. centric core of participation’<sup>75</sup> on *YouTube* evidenced in Lange (2007b), and, more significantly, the locally-oriented character

---

<sup>74</sup> For censorship reasons, *YouTube* access is banned in some countries, like China, Iran, Morocco, Thailand and Turkey. Yet due to the popularity of the Website and its influential business power, the list of banning countries varies from time to time (in August 2008, Thailand, for example, had provisionally suspended the ban) and this may explain while some of the ones just mentioned are nonetheless present in Fig. 21 (unless, of course, the video’s proposition is false, which can not be ascertained without verifying it against the (You)Tuber’s offline location).

<sup>75</sup> US does not only predominate as the location from where to (You)Tube, but it also seems to have a peculiar central position in the (You)Tubers’ system of reference; indeed, unlike *ChangeDaChannel*, many US (You)Tubers merely mention the state abbreviation (e.g., ‘CA’ for California, ‘OK’ for Oklahoma) without evidently feeling the need to mention the country. In turn, non-US (You)Tubers frequently locate their country in a wider context, either verbally (e.g., ‘Croatia, Europe’) or by means of other semiotic resources, i.e., by showing a map representing the neighbouring countries or zooming on the country by means of software tools like *Google Earth*.

of video-interaction observed by Benevenuto et al. (2008a)<sup>76</sup>. Indeed the most frequent responded country is also the initial video's.

Country	Vid.	Country	Vid.	Country	Vid.	Country	Vid.
USA	287	Argentina	5	Venezuela	3	Cyprus	1
UK	102	New Zealand	5	Egypt	2	Estonia	1
Canada	65	Switzerland	5	Hungary	2	Iceland	1
Germany	31	Belgium	4	Israel	2	India	1
Australia	23	China (Hong Kong)	4	Japan	2	Kuwait	1
Italy	20	Colombia	4	Lithuania	2	Morocco	1
Mexico	19	Denmark	4	Philippines	2	Peru	1
Spain	13	Ireland	4	Puerto Rico	2	Samoa	1
Brazil	11	Portugal	4	Saudi Arabia	2	Slovakia	1
France	11	Romania	4	Taiwan	2	Slovenia	1
The Netherlands	9	Austria	3	Thailand	2	South Africa	1
Sweden	7	Chile	3	Turkey	2	South Korea	1
Finland	6	Croatia	3	Azerbaijan	1	Ukraine	1
Norway	6	Czech Republic	3	Bangladesh	1	Yugoslavia <sup>77</sup>	1
Poland	6	Honduras	3	Costa Rica	1		
<b>Total</b>							<b>716</b>

Fig. 21. Countries represented in the video-thread<sup>78</sup>.

Nonetheless, the significant quota of non-US responses is favoured by (1) the topic, which prompts an international participation (as in international introductions, *ChangeDaChannel*'s anticipated answer mentions the country 'USA', along with the State), and (2) the use of English, the international language *par excellence* (cf. Crystal, 1997). The mode chosen for verbal language also contributes to the international participation; indeed, English used in writing rules out the problems of accent and fluency of the spoken mode. This fact is evidenced by comparing this trend to the more locally-oriented composition of the 'Why Do You Tube?' video-thread, whose initial video employs speech, and only 3 responses out of 356 feature a non-native English speaker (Adami, forth.).

Totally considered, the international reach of the locations represented in the thread evidences to the fact that participants in video-interaction share representational practices and resources which go beyond their national specific ones to make meaning, like their mother tongue or other culture-specific signs. Indeed, they are

<sup>76</sup> Their statistical analysis of 3.4 million video responses shows that '40% of the responded videos have a percentage of responses from the same country superior to 60%' (2008a: 1).

<sup>77</sup> The video uses this old political label; it is maintained here, since the actual country is not recoverable from any clues.

<sup>78</sup> The locations outnumber the videos (716 against 708); 8 videos represent two countries at once, i.e., in the case of 'nomadic' (You)Tubers, who declare to (You)Tube sometimes here and sometimes there, specifically in: (1) Canada, USA; (2) Egypt, Saudi Arabia; (3) Finland, Germany; (4) Hong Kong, USA; (5) Spain, Argentina; (6) Spain, Yugoslavia; (7) UK, Germany; (8) USA, Venezuela.

well aware of the international nature of the interaction and adapt their selection of semiotic resources accordingly (cf. Adami, forth., for an analysis of internationally known stereotyped used to represent the country in the responses).

By taking up *ChangeDaChannel*'s content-specific prompt, these 708 responses provide the prompted geographical information and thus interpret *ChangeDaChannel*'s request and the interactional exchange *ad litteram*, i.e., as a request of information.

### 3.1.2 The non-geographical location

The remaining 36 videos ignore *ChangeDaChannel*'s content-specific prompt and answer the topic question without mentioning any country. Thus their authors can (You)Tube from some places in the outer space (e.g., *Venus*, *Uranus*, from *Planet Mars*, from *a galaxy far far away*), from a located but unnamed land (*six thousand kilometres*, *six time zones and an ocean away from New York*), from qualified but unidentified lands (*a world full of despair and hurt*, from *a lonesome valley*), from fictional places (*Hogwarts*), from some domestic ones (*my wardrobe*, *my special room*, *my house/home*, *my college dormroom*, *inside of my blanket fort*, *inside my lamp*, *in bed*, from *the toilet*, from *your closet*, *@work*), from some provocative ones (*your mom*), or from some apparently nonsensical ones like *a beer train*.

All these very diversified locations are here quoted in their verbalized representations; however, like the ones which represent a geographical location, also these video responses frequently use images (Fig. 22), such as photos of the planet, filming of the (You)Tuber wearing an alien-like mask, of a hand carrying a written paper from inside a lamp, of the (You)Tuber getting out from the closet with the handwritten location in her hands, or of a toy-train with carriages branded with beer names, filmed running on the railway getting out from a vendor machine. In this latter case, the topic answer seems just the occasion to show the (You)Tubers' handicraft; this is a particularly evident case in which the request is taken as the prompt for an interest-driven response.



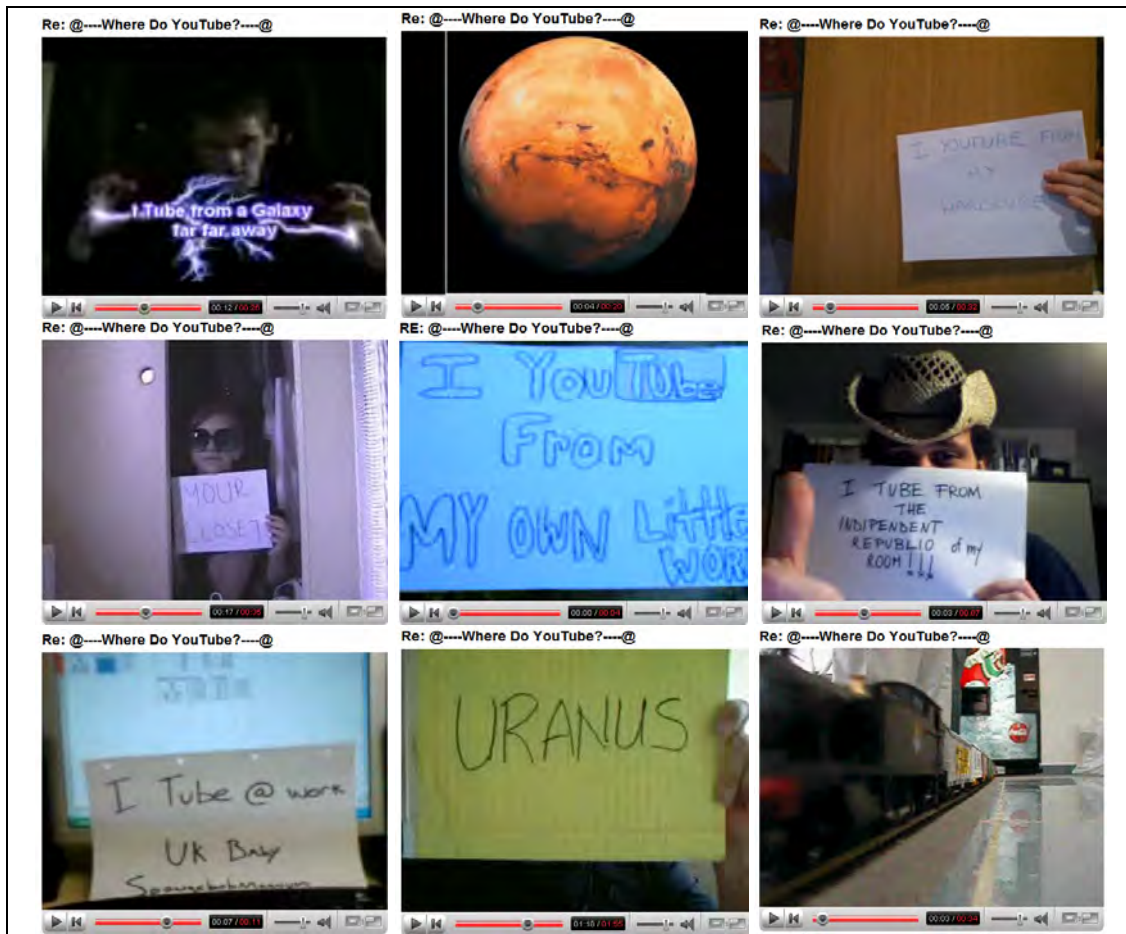


Fig. 22. Non-geographical locations represented through various resources in the responses.

By representing the location by a combination of resources, these videos answer the question by providing unexpected content, so as to produce humour. Reasonably, these respondents know exactly the expected answer (i.e., the geographical prompt of the initial video), but choose to exploit maximally the ambiguity of the topic – the meaning space left open to interpretation – so as to reply still ‘on topic’, but in an unconventional way, which breaks Gricean cooperative principles (and thus produces humour). This way they simultaneously (a) avoid any disclosure of potentially dangerous information concerning their location (mistrusting *ChangeDaChannel*’s encouraging prompt) and (b) show their interested interpretation of the request and of the interactional exchange in terms of a playful provision of topic-related entertaining content rather than of (serious and reliable) information.

### 3.1.3 The prompt-response continuum of the represented location

Consciously misinterpreting the question, these playfully topic-subverting videos can be placed towards the end of a prompt-response continuum, on which responses can be arranged according to how the question is answered. At the one end are the

responses which take up the request *ad litteram* and provide the prompted geographical information as answer. At the opposite end are the videos which explicitly challenge *ChangeDaChannel*'s request. One shows a handwritten insult ('WELL RNT U A NOSEY BASTARD') instead of a location, while three others represent a provocative non-geographical location ('[I TUBE FROM] YOUR MOM!!!')<sup>79</sup>; cf. Fig. 23:



Fig. 23 Insults and provocative non-geographical locations.

Three responses 'spoof' *ChangeDaChannel*'s video (Fig. 24). One rephrases *his* typed introduction:

a subscriber said I need more subscribers...  
so I decided to do this video to get more subscribers

By so doing, this response expresses its interested interpretation of the initiator's purposes in posting his video request, i.e., to get more subscribers. Another response rephrases the whole content of the initial video and reproduces its textual organization. The response is shot in black and white with a heavy metal soundtrack in the background; on a black screen the following (white) typing appears:

A tuber said i wank too much in my vids so I jerked off just before

Then the (You)Tuber appears showing a handwritten paper:

I TUBE FROM THE TOILET

He turns the writing towards him, reads it, shakes his head, turns the page and starts to write; he then shows this second writing:

I SHIT ON ~~FROM~~ TOILET

A thumb-up gesture precedes the black screen with the typed teasing closing:

<sup>79</sup> Beside its sexual connotation in the context of 'I tube from your mum', the phrase is used to nullify the significance of the interlocutor's request and to avoid a cooperative reply.

still here?  
ok your problem i need to jack off



Fig. 24 Responses parodying the initial video.

A third response parodies *ChangeDaChannel*'s video description and the motivation for his no-talking rule:

\*Write down or say or show where YouTube from:)Ill leave this up for a couple-few weeks.\* A subscriber PM'd me and said that I talk too much in my videos...so i decided to do a silent one and too also see where you tube from:) Ill give it some time and then do a sequence from the RESPONSES. Thanks again for all your support....EVEN though i talk too much;) (more) (less).

Somewhere in the middle of the continuum are 21 responses which provide initially an unexpected answer, but represent also the geographical location. By answering with cinematographic references – e.g., ‘I YOUTUBE FROM THE DEATH STAR’ overlaid on a Darth Vader-masked (You)Tuber (cf. Fig. 25.a) – or implying fictional crime-plots – e.g., ‘fm my captors basement’ on the shot of a guy sat on a chair, his mouth sealed with duct-tape, hands and feet tied (cf. Fig. 25.b) – these (You)Tubers enact their unique performance so as to produce humour and further differentiate themselves, before representing the geographical location. In so doing, they initially ‘frame’ (Bateson, 1972) their interpretation of the exchange as playful, eventually complying with *ChangeDaChannel*'s content-specific prompt, perhaps hoping that this differentiation-within-compliance may gain their inclusion in the summary.



Fig. 25 Enacted non-geographical locations.

Other videos comply formally with the content-specific prompt without providing any substantial information. Indeed, as anticipated, in two videos the geographical location is not recoverable: one films an unreadable handwritten paper; the other is a very fast *Google Earth* zoom from Earth to an unidentifiable block of houses.

Both responses exploit technical affordances of the medium – *YouTube* videos’ low-definition<sup>80</sup> for the former; the absence of a slow motion option for the latter – to prevent the identification of the location and, by so doing, they find a way to answer the question as prompted, without disclosing any information. The concealing of information is maximum in these responses; indeed, in both, the (You)Tuber’s identity is unrecoverable too: in the first the face details are concealed by means of colour effects, while the second response does not represent the (You)Tuber at all. In other terms, these responses are maximally evasive, while simultaneously being ‘perfectly’ topic-related and even geographically ‘themed’ (i.e., they find a way of answering the question on topic as prompted, without however providing a substantial answer).

Rather than providing an answer, 14 responses<sup>81</sup> ask the viewers to guess the geographical location on the basis of some clues. A response asks viewers to guess the location by featuring various shots of a town. Another one is structured as an home-made holiday-video (which shows people having fun at a seaside location) preceded and followed by a typed invitation to guess the location. The video does not give any specific clues to identify the place (either visual, like flags or maps, or linguistic, by mentioning any geographical name), nor does the para-text. Furthermore no feedback is given to the various comments which try to guess the country (the tentative answers are Costa Rica, Nicaragua, Ukraine, Brazil, Ecuador, Australia, Tasmania, New Zealand, Indonesia, and Thailand). In other terms, the topic question remains without a definite answer; it is thus an instance of topic-related and geographically themed ambiguous answer.

Another response shows photos of monuments and the outline of the country map (Fig. 26.a.), asking in typing viewers to recognize the country (which is correctly guessed in the comment section, i.e., Portugal, as confirmed by the respondent’s feedback). In a further response the (You)Tuber carries a paper with a printed symbol of lily and the question: ‘do you know what it is?’ (Fig. 26.b.). In this case, only the ones who have the needed specific-knowledge of the location can recover the information. Comments attempt variously at guessing (‘Florence’, ‘Le Fleur du Rois – France’ and ‘Louisiana’) but the (You)Tuber does not give any feedback to them (eventually, on her channel, her profile declares she is settled in Louisiana).

---

<sup>80</sup> A later introduced option enables uploaders to select a high-definition quality for their videos; yet, maybe not incidentally, the here-discussed video response still cannot be played in high-definition.

<sup>81</sup> All locations but one – i.e., the here-below discussed holiday-video – have been retrieved, by means of unambiguous clues given in the responses or in the paratext. Therefore 13 of these 14 responses are included in the list of locations of Fig. 21.

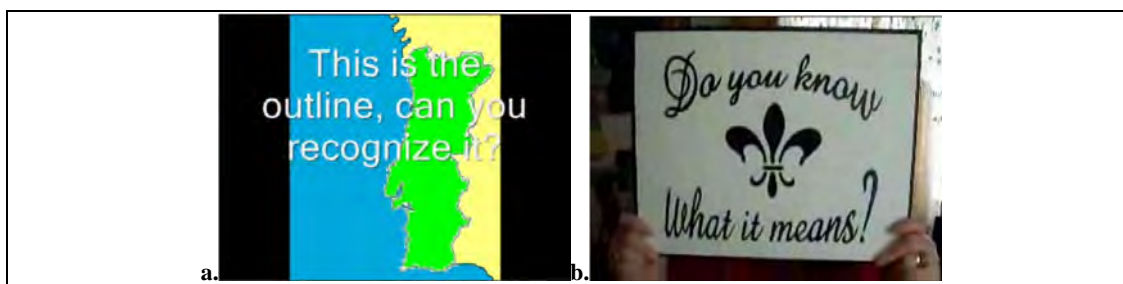


Fig. 26 Guess-type responses.

Following the geographical theme but eliciting the answer, rather than providing it, these (You)Tubers take *ChangeDaChannel*'s content-specific prompt as a chance to perform their (quiz-game) show and entertain the viewers, challenging them to guess. They also show their interested understanding of the interactional exchange as a themed performance, rather than as a provision of information.

At the very far *ad litteram* end of the continuum are videos which disclose more personal information than prompted; indeed, besides (or instead of) representing the country, 72 responses focus on domestic locations. They can zoom-in from the country to shots of the (You)Tuber's room/house, generally through visual resources, by combining zoomed shots of *Google Earth* with some final shots of the house, sometimes accompanied with written labels functioning as 'captions' (Fig. 27).

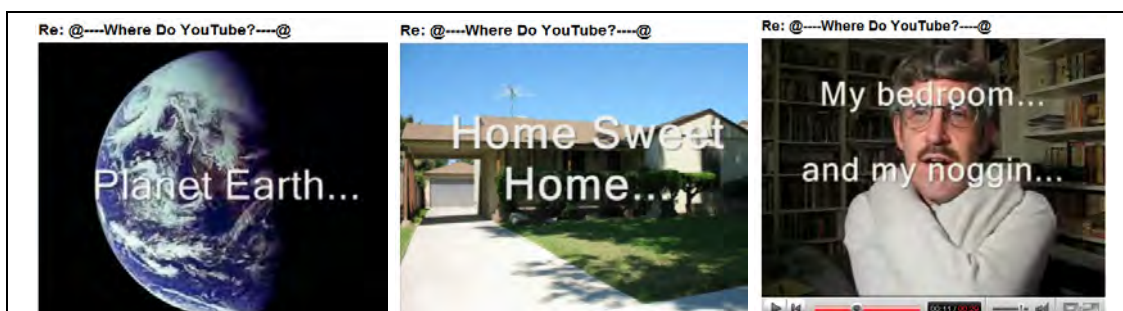


Fig. 27 Screenshots from a zooming-in response.

Alternatively, they can de-zoom from there up to the country; this is done more often with linguistic means, as in a video where the (you)Tuber shows the following handwritten papers:

I TUBE FROM  
MY CHAIR WHICH IS..  
IN MY ROOM WHICH IS..  
IN MY APARTMENT WHICH IS...  
THE TOWN OF ST-JENV-SUR-RICHELIEU IN...  
QUEBEC CANADA

Some (You)Tubers introduce the viewers to a 'virtual' tour of their room or house



(by turning the camera around the place or by means of a slideshow of photos); they can even show their neighbourhood (by shooting the view from their window) and some (fewer) others give information on their personal network, by filming or showing photos of themselves, their relatives and friends (Fig. 28). In this way they disclose even more personal information than prompted in the initial video.



Fig. 28 Private environment as the represented locations.

By giving access to their personal sphere, these videos construct a relationship with the viewers which is closer than the prompted introductory one. They again exploit the ambiguity of the topic and actualize one of the possibilities of the prompt, by answering the question ‘where do you tube?’ in its maximum *ad litteram* sense (i.e., the exact spot from where they (You)Tube), thus taking the chance to show something of their lives. In other words, these respondents select the criterial aspects of the prompting video according to their interest, so as to represent their domestic identity rather than (or besides) their nationality.

### 3.1.4 The signifiers of the location

As shown in Fig. 29, responses represent the location through a wide range of resources. Most of them take up *ChangeDaChannel*’s form-specific prompt; indeed, the location is most frequently written (678 videos), and, particularly, handwritten (519). In turn, while 88 videos represent it in speech, only 41 answer the topic question *exclusively* through speech (with no other representations of the location); significantly, two of these are further reposted with a handwritten reformulation.

In sum, most video responses not only take up the prompted (geographical) content, but also ‘keep to the theme’ (as stated in one of them). All these (hand)written responses are consistent with the prompted (hand)written mode, and, even more, with *ChangeDaChannel*’s ‘no-talking rule’, signified both through embodied and disembodied resources (his mouth sealed; ‘so I decided to do this without talking’).

Type of perception	Mode used to represent the location	Videos
Visual	Hand-writing	519
	Type-writing	159
	Still image	89
	Filming	68
	Map	61
	Flag	33
	Satellite shot ( <i>Google Earth</i> )	25
	Language	23
	Drawing	23
	Emblem	15
	Clothing	15
	Road sign	6
	Globe	3
Auditory	Speaking	88
	Soundtrack	40
	Singing	3
<b>Total</b>		<b>1170</b>

**Fig. 29. Modes used to represent the main<sup>82</sup> location.**

22 responses meta-reflect on *ChangeDaChannel*'s prompt; they can do it in writing (e.g., a handwritten 'Just like the others. I won't say a word. Respect!'), with gestures (e.g., 'zipping' one's mouth shut by passing the first and index fingers from one side to the other of the mouth), or by filming the (You)Tuber with duct-tape sealing her mouth (significantly, one has the location written on it; cf. Fig. 30). A video salutes the 'no-talking rule' ('Yay! A video where I don't need to talk!'; handwritten excerpt), while others highlight the difficulty of complying with it; cf. for example, 'I can't take it I can't not to talk in a video I'm sorry', uttered at the end of a video where the (You)Tuber has handwritten the location. Finally, meta-statements are also present in the paratext, like the following comment given by the (You)Tuber to another one asking 'why u no talkie?':

oh that's because the person i was responding too's video was silent it was kind of a thing for that vid...watch it and you'll see...ttyl

<sup>82</sup> 'Main' refers to all *foregrounded* representations of the location. All background information on the place of (You)Tubing in videos (e.g., the room) has not been considered, unless focused by some deictic reference (such as pointing gestures or verbalized resources like 'this is where I tube').

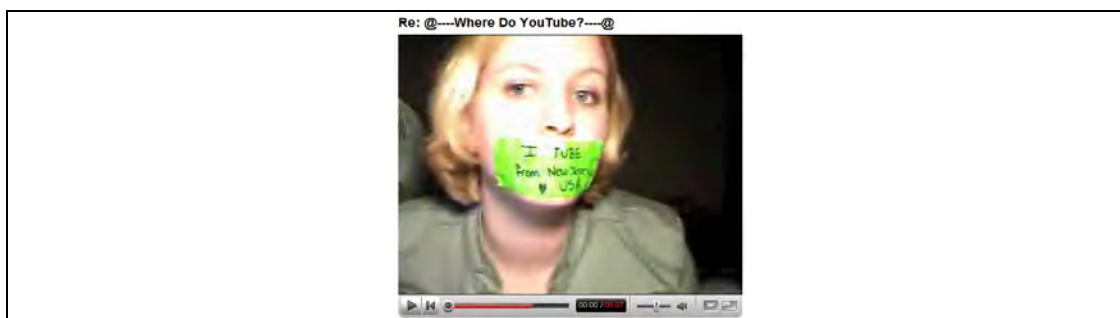


Fig. 30 Duct tape as signifier of compliance with the prompted ‘no talking rule’.

Definitely, prompted by the initial video, the mode of writing plays a major role in the thread. Indeed, excluding from the count any possible background information shown through the filming of the (You)Tuber, 458 videos represent the location by means of only one mode. 396 of these employ either hand- or type-writing, while only 62 select a mode other than writing: 47 employ speaking; two sing the name of the location; nine represent the location exclusively by filming it<sup>83</sup>; three use *Google Earth* software; one uses the location emblem. When it is not verbalized, the location is mainly represented through more than one mode. Without verbalization, the location is hardly identifiable within the geography discourse (which is the one that is both prompted by the initial video – cf. ‘I TUBE FROM CALI, USA’ – and on which most responses draw), unless highly ‘geographical’ means of representation are used, e.g., *Google Earth* software or the image of a map, or widely known stereotyped symbols of the location are used, e.g., the image of the Tour Eiffel for Paris/France. Conversely, when geographical resources appear (e.g., the map and/or the flag or the country name), still and dynamic images of the location function as a further specification and, especially when they include the (You)Tuber’s persona – as in the case of a (You)Tuber filmed while dancing in front of Sydney Opera House – they are chiefly understood as signs of truthfulness and authenticity, which is not a secondary issue in online communication.

As Fig. 29 shows, beside the general attuning with *ChangeDaChannel*’s prompted mode, the video-thread attests also an impressive wide range of multimodal resources for the representation of the location (1170), which outnumber by far the responses (746). Apart from or along with writing, the location is represented through still/dynamic images of the country or of well-known national symbols (e.g., the *Guinness* beer logo for Ireland, a photo of Valentino Rossi for Italy); it can be indicated (by pointing to it) on a globe or represented through clothes (e.g., a hat of the local baseball team), maps, flags and satellite shots (cf. Fig. 31).

<sup>83</sup> In these latter cases, the exact town/country name could be retrieved only from the para-text, while the video itself did not supply (the researcher with) enough clues to identify it.





Fig. 31 Signifiers for the location.

When verbalized, the location can be sung rather than spoken and, when written, it can appear on a road sign or within the town/country emblem. Also stereotyped phrases of the native language<sup>84</sup> contribute to represent the location, like ‘ole!’ for Spain and ‘Aloha’ for Hawaii (cf. Adami, forth., for language as signifier of the location in the video-thread). Finally, 40 videos use a topic-related soundtrack, like the national anthem, a song mentioning the town (e.g., ‘Barcelona’, by F. Mercury and M. Caballe), a stereotyped national genre (e.g., bagpipes sounds for Scotland), or a famous national singer’s tune (like Edit Piaf’s for France).

The signs used to represent the country vary from the ones which rely on the discourse of geography (the country map, its satellite shot, name or flag) to the ones which rely on the discourse of national identity (by using internationally known stereotyped emblems, which in this context function as signifiers of the country, such as the anticipated logo of the *Guinness* beer for Ireland, the t-shirt of the national football team for Brazil, or photos of national celebrities, i.e., Michael Schumacher

<sup>84</sup> Language here consists of idiomatic phrases used as clues for the location of (You)Tubing. So, the accent has not been considered, being a clue of the (You)Tuber’s origins rather than of the location. English is not counted, functioning as an international language in the thread; it is included only in the case of six Canadian videos; indeed, five of these represent a written interjection (‘eh?’) stereotypically attributed to Canadian English, its stereotype status is further confirmed by its rejection in the sixth video from Canada: ‘and we don’t say eh, ok?’ (handwritten excerpt).

for Germany). In turn, as seen, other responses use actual shots of their local environment, thus providing new information on their place. Also the soundtrack can be either a widely known tune, institutionally or stereotypically associated with the country, or a less famous song, authored by someone coming from the represented country and credited in the video, so that it provides new information on the country.

In sum, at an ideational level, the signifiers of the locations either give new information on the (You)Tuber's environment, by showing the (You)Tuber's specific (domestic) location, or rely on some shared knowledge and draw on the discourse of geography (by using its conventional resources, like maps), or of national identity, by using stereotyped signs which symbolize the country, so informing on the (You)Tuber's nationality.

In all cases, it becomes apparent that many videos do more than 'merely' responding to what prompted, not only in supplying more content, but also in using a wider range of resources. This is a formal way to present one's own distinctiveness – and thus to construct identity – within a semiotic space where creativeness is highly valued. Rather than merely providing information, any semiotic act plays an 'interpersonal function' (Halliday, 1978) in the thread. The specific prompt-response relation instantiated in each video communicates its author's identity and relationship with the interactants, hence attuning with the initial video signifies solidarity in the interaction, while variation (within or beyond attuning) signifies 'I am different'.

In other terms, responding to a video request becomes a creative challenge within a set of limited possibilities; it is a matter of producing the maximum extent of variation stemming from a given 'kernel'. Ultimately, the great representational variation is the (You)Tubers' response to a prompted challenge implied in *ChangeDaChannel*'s request, which can be verbalized with something like 'show me how you can answer differently', and consequently, 'how you *are* different'. A video description discusses this challenge:

Felt like a challenge, saw this asking for responses and thought it'd be fun. Decided to make it themed, first considered a slideshow (fancy-edited of course) of pics from my town/etc, but then captured a few random weather clips on the morning news.

Another description (of a video which employs *Google Earth* software) highlights the high-rated value of originality in videos:

Okay, so I wasn't the only one who had this particular idea nor was this one the best out of all the responses that were along the same lines BUT I think that I'm a close second and that's good enough for me! ;)'

As evidenced also by this hedging description (i.e., which backs up the (You)Tuber's creation, by acknowledging its lack of originality), an implied 'creative' performance is generally understood in the practice of (You)Tubing. (You)Tubers commonly exploit the potentialities of the medium and prove to be different, witty, and creative,

so as to surprise the viewers, catch their attention, and entertain them. The reward to their effort is well quantified by the Website's public indexes, like the video's number of views, comments and achieved rating. This generally understood challenge is specifically reinforced in video-interaction, where the room for differentiation stems from a given theme/form set by the initial video, so that the constraints within which creativity can be enacted make the task even more challenging. Even more in this thread, the challenge for differentiation is further reinforced through the prospective reward of an inclusion in the video-summary. Understandably, the more the responses, the more they need to differentiate themselves to gain inclusion; simultaneously, being the selection (both within the video-thread and in the summary) subjectively done by the initiator, (You)Tubers need to carefully balance differentiation and compliance with the editor's interests to minimize the risk of exclusion.

### 3.2 Formal relatedness: The (marked) mode as attuning device

Apart from the answer to the topic question, another element which contributes to establish relatedness is the formal correspondence of the response to the initial video, in terms of multimodal deployment. As anticipated, the consistency in mode functions as an 'attuning' device, because it enables the respondent to attune with the 'register' of the initial video (the mode is indeed one of the components of the register, according to Halliday and Hasan, 1976). In the thread, the most frequent type of response is an indoor shot of the (You)Tuber facing the Webcam, carrying a sheet with a handwritten geographical name<sup>85</sup>, thus responding to the request by deploying its very same multimodal pattern. Fig. 32 summarizes the multimodal deployment in the 748<sup>86</sup> videos that answer *ChangeDaChannel*'s question. As can be seen, most video responses are highly multimodal, generally deploying some writing (704 videos), dynamic (670)<sup>87</sup> and/or still (118) images, and/or drawings 160; in most cases, they also employ also some auditory modes, like a soundtrack (468 videos), environmental noises (301) and/or something spoken (113) or sung (6) by the (You)Tuber.

---

<sup>85</sup> This very frequent type of responses gives simultaneously two types of information related to the topic question, i.e., the geographical one by means of handwritten language and the specific-domestic one represented by the elements of the room shown in the background.

<sup>86</sup> To the total of related video responses (746) considered for the analysis of the location, here two duplicates have been added, in two cases where a spoken video response has been re-posted by the same (You)Tuber with a written reformulation; hence, in both cases the second posting, which has been considered as 'duplicate' for the analysis of the represented locations (because it provides the same information twice), is to be considered as a different video in the analysis of the multimodal deployment in videos (even more, as discussed, this change in the mode is highly significant, since it attunes the response with the handwritten mode of the initial video).

<sup>87</sup> The labels 'dynamic' and 'still' images are used here very specifically. In fact, every resource can be said to be 'dynamic' in videos, since it displays through time. Here, the label 'dynamic images' does not include slideshows of still images, assuming that the watcher's eye perceives them as a sequence of stills. Still images only refer to photos, this way further distinguishing drawings.

Perception	Mode	Videos	Sub-mode	Videos
Visual	Writing	705	Hand-writing	539
			Type-writing	323
	Dynamic images	669	Filming	641
			Animation	19
			Software-generated	25
Auditory	Still images	279	Drawings	161
			Photos	118
	Black and white	57		
	Speaking	113		
	Soundtrack	467		
	Singing	6		
	Noise	301		

**Fig. 32. Multimodal deployment in the video responses**

Interestingly enough, the most frequent modal resources in the video responses are the ones occurring also in the initial video, i.e., handwriting (539), filming (641), soundtrack (467) and type-writing (323). Yet only 57 videos are filmed in black and white; this could be a clue that, unlike individual texts (Kress and van Leeuwen, 2002), interaction does not use colour as a framing resource; in our terms, colour is an intra-textual cohesive resource rather than an inter-textual attuning one.

Type of perception	Mode used to represent the location	Videos
Visual	Hand-writing	519
	Type-writing	159
	Still image	89
	Filming	68
	Map	61
	Flag	33
	Satellite shot ( <i>Google Earth</i> )	25
	Language	23
	Drawing	23
	Emblem	15
	Clothing	15
	Road sign	6
	Globe	3
Auditory	Speaking	88
	Soundtrack	40
	Singing	3
<b>Total</b>		<b>1170</b>

**Fig. 33. Modes used to represent the main location.**

At this point, it is useful to compare the general multimodal deployment in the thread detailed in Fig. 32 with the modes used to represent the location discussed in 3.1.4. To this aim, the related table listing the representational resources of the location is re-presented in Fig. 33.

By comparing the two sets of data in Fig. 32 and Fig. 33, the general trend of consistency in the use of modes with the initial video is even more evident. While, as indicated in Fig. 32, some videos (157)<sup>88</sup> display both hand- and typewriting, the location is far more frequently handwritten (519, as detailed in Fig. 33) than typed. This is consistent with *ChangeDaChannel*'s initial video, which uses typewriting only for introductory and conclusive messages, while devoting handwriting to the topic question and his anticipated answer. Furthermore, out of the 641 videos displaying some filming (Fig. 32), only 68 use it to represent the location (Fig. 33). This general trend is also consistent with the initial video, which filmed *ChangeDaChannel* at a mid-close shot, but did not use filming to represent the location (to this regard, 588 videos show the (You)Tuber's face at the camera). The same is valid for the soundtrack, which is used in 467 videos, as is in *ChangeDaChannel*'s, while only in 40 it represents the location.

The analogous trend observed on the side-reference thread used in the pilot study of the research (cf. Chapter 3, Section 4.2) evidences a 'tuning' function played by the mode prompted in the initial video. Indeed, as discussed in Adami (2009a), in the case of the 'Why Do You Tube?' thread, the initial video films the (You)Tuber *speaking* his reasons for (You)Tubing and asking *orally* his viewers to post their reasons. This time, the great majority of video responses employ spoken language in answering. More specifically, excluding the 23 inaccessible responses (i.e., set to private), 311 out of 333 responses make use of spoken language, while only 39 videos have some written language (alone or along with speech), and further 35 videos employ other visual resources, either dynamic or still images.

In sum, in both video-threads (totally amounting to more than 1,000 videos), the selection of the main mode for replying is prompted by the initial video, as represented in Fig. 34. Hence the mode, prompted by the initial video, functions as a tuning device in video-interaction.

---

<sup>88</sup> This datum is given by the sum of handwriting and typewriting in videos (539+323=862) subtracted the total number of videos which display some writing (705).

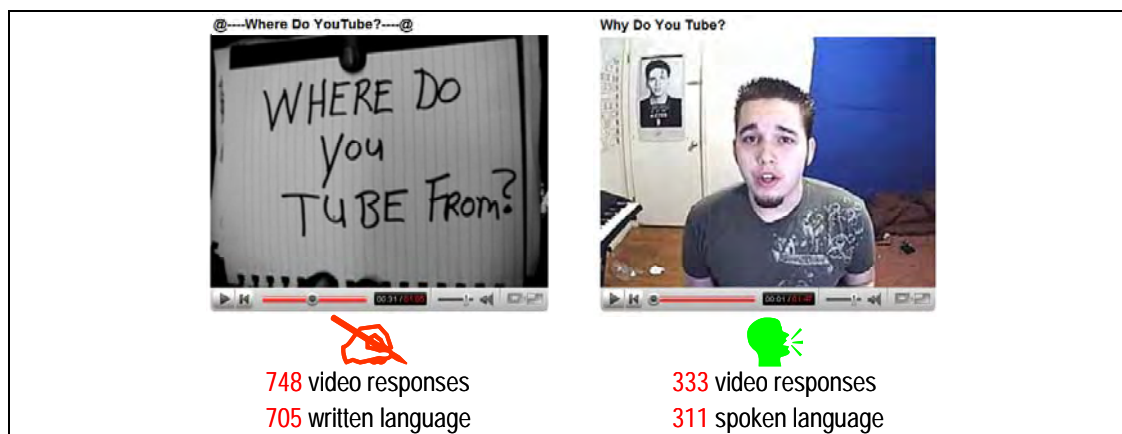


Fig. 34. Attuning in mode in the two video-threads: ‘Where do YouTube?’ and ‘Why Do You Tube?’.

The attuning in form is by no means exclusive to video-interaction; in this regard, cf. Burgoon et al.:

Interactants tend to entrain and coordinate their interaction with the nonverbal and verbal behaviors of a co-interactant, producing a prevailing pattern of behavioral matching, reciprocity, and synchrony, even during deception (1999: 671).

However, in video-interaction, which enables so many semiotic modes to be deployed, the mode becomes an essential element of the ‘theme’ to be maintained<sup>89</sup>, within which the maximum extent of variation is produced.

At a finer-grained analysis, each video is unique, in terms of soundtracks (or environmental noises), colour palettes and effects, (You)Tubers’ personae (e.g., human faces, animated avatars), facial expressions and gestures, handwriting styles, writing colours, fonts, layouts, supports and materials, camera angles and positions (so that handwriting is shown to a fix Webcam or the camera moves and films it), backgrounds behind (You)Tubers, them enacting the process of writing or showing just its final product, and so on. If semiotic resources are always transformed when meaning-makers use them (Bezemer and Kress, 2008), this is maximally true in video-interaction, where differentiation and ‘originality’ are highly valued.

Therefore, when deploying the same multimodal pattern, video responses maximally ‘tune in’ with the initial video simultaneously showing their distinctiveness. In this sense, the thread is akin – yet multi-authored and multimodal – to well established genres in music, such as ‘variation’, which dates back to the Ancient Greeks and is

<sup>89</sup> One must not underestimate the relation between the topic and the chosen mode of representation; indeed, while the topic question ‘where do you tube?’ may be answered more straightforwardly through different modes, replying to the thread ‘why do you tube?’, i.e. representing the reasons for (You)Tubing, may be more easily done in speech rather than – say – through images; in other words, the affordances of each mode might influence the choice of the mode used in videos. Most likely, all these factors contribute to the respondent’s selection of the mode, driven both by her interested interpretation of the prompt and by the most apt form available to her in that context.

epitomized by, e.g., Bach's *Goldberg Variations*. In turn, the multi-authority and multimodality of this attuned-and-varied performance can easily find its counterpart in contemporary forms of improvisation, such as jazz sessions or free-style dance and rap (or hip hop) collective performances. Given that video-interaction is a very recent practice, conventions for both attuning and variation are not already fixed (unlike for classical or jazz music, or even for free-style and hip hop performances, which are older pop-culture practices than video-interaction). Yet the practice is analogous: an initiating move sets a theme, a pattern is identified by the interactants and is responded by both attuning with it and deploying the maximum extent of variation, in content and in form, following the participants' interests (and available resources).

In video-interaction this formal attuning is so significant that, when topic-relatedness is absent, the attuned multimodal pattern functions as a formal clue of relatedness (Fig. 35), as in the video featuring the (You)Tuber showing a handwritten insult instead of the location ('WELL RNT U A NOSEY BASTARD'). Analogously, another response shows what, only by virtue of this multimodal attuning, can be interpreted as an evaluation of the initial video: a handwritten 'cool! I KNOW' (instead of the location). Form becomes content when the interaction is taken as a chance to perform a 'unique-but-attuned' performance.

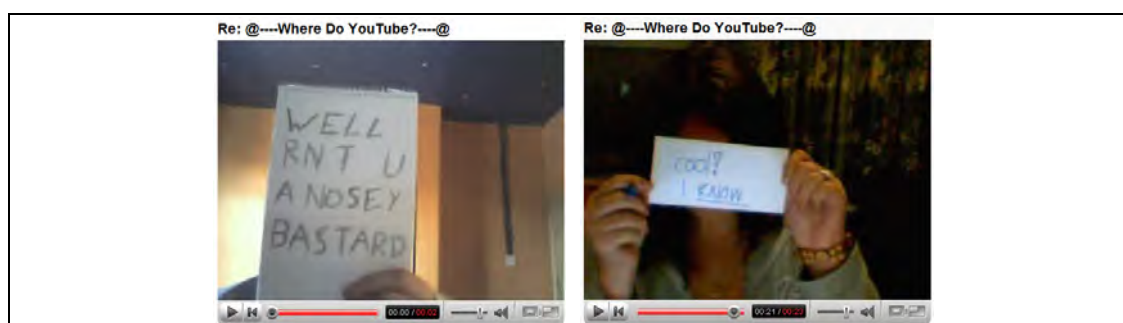


Fig. 35. The attuned mode as the only clue of relatedness.

True, one cannot consider a correspondence at the level of modes as a 'cohesive' tie, because modal realizations are essentially formal, while, in Halliday's terms, cohesion is essentially semantic and involves semantic ties. Nevertheless, especially when the mode selected for the topic is a marked one, this attuning device constitutes a strong clue for relatedness (even if it constructs relatedness in form rather than in 'meaning'<sup>90</sup>). Therefore, either we consider this attuning device as a 'special type' of cohesive tie, or more wisely (so that we do not stretch Halliday's notion outside its semantic domain), we can conclude that cohesion is not essential to relatedness when formal 'attuning' is there, as evidenced by these attuned (but out of topic) responses.

More specifically, in the 'Where Do YouTube' thread, the mode used to represent

<sup>90</sup> Note that form has always meaning – or, better, that formal elements can be associated with criterial aspects of meaning (e.g., the meanings which can be produced by the type fonts in printed material).

the topic question is a marked one; indeed, handwriting on paper is an unusual modal realization, for videoblogs in particular (which generally use speech for their ‘ranting’), and for online-videos in general (which usually select the form of writing which is more readily available in that medium, i.e., typing on the screen). Its markedness makes the semiotic mode a rather salient element in the video so that it prompts attuning and can further function as a clue for relatedness. Therefore, a handwritten representation is not only more ‘attuned’ with the initial video, but, even more, when all cohesive (semantic) ties are missing, it functions as the only clue of relatedness. So, a response which shows a handwritten paper moved towards the Webcam is clearly understood as a related response to the answer even if it does not represent any location, but – as anticipated earlier – an insult (‘WELL RNT U A NOSEY BASTARD’) or an evaluation (‘cool!’). Significantly, in watching these responses, the referential system for interpreting the verbal expressions as an insult (for the former) and as an evaluation (for the latter) is established only by virtue of the attuned mode. Indeed, the ‘you’ is interpreted as referred to *ChangeDaChannel* right because the form (handwritten paper) of the representation perfectly matches the one chosen by *ChangeDaChannel* in his video; analogously, the scope of ‘cool’ is referred to the initial video because it is handwritten on paper. In sum, in presence of an attuned marked mode, responses are related no matter their content.

### 3.3 Textual organization: The relatedness-continuum

On the basis of how the representation is textually organized, responses can be arranged on a relatedness-continuum, from the ones which display more clues and hence the greatest degree of relatedness, to the ones which have almost no clues (and hence are more likely to be perceived as less related if not even totally unrelated).

This relatedness continuum is constructed by the extent to which elements are present and on their textual organization in the responses. The more redundant<sup>91</sup> their presence and the less marked their textual organization, the more the perceived relatedness with the initial video. Hence, as analysed here-after, responses are maximally related when they recap the whole exchange (3.3.1), then come the responses which represent the location in an unmarked organization for the answer to the topic-question (i.e., [Given: agent + action]; New: location; cf. 3.3.2), then the ones which represent the location in a marked position for an answer, i.e., as Given or as circumstantial element (3.3.3). Paratextual clues in the video thumbnail (a), in the title (b), in the description (c), and, in a different way, in the comment section (d) also contribute to modulate the extent of relatedness (3.3.4).

---

<sup>91</sup> Redundancy is a means of foregrounding a sign. It is a means of producing explicitness. The least redundant way of answering the question is showing the (You)Tuber filmed sat at her pc (cf. 3.3.5).



### 3.3.1 Narrative structure

Video responses can be structured as either a retrospective or a prospective narration of the interactional exchange (or both).

#### *a. Retrospective narration of the exchange*

Retrospectively, responses can start by citing (either with images or with language) the initial video and the action of responding to it (73). They can also explicitly mention *ChangeDaChannel* (38), and/or the topic question or rephrase his request (185), e.g., ‘you want to know where I tube from?’. Some declare the function of the video at its very end: (‘this is a response to...’ or ‘this goes out to...’), as when writing the recipient’s address on an envelope. A response represents the location on a post it attached to the pc screen where the initial video is playing (Fig. 36).

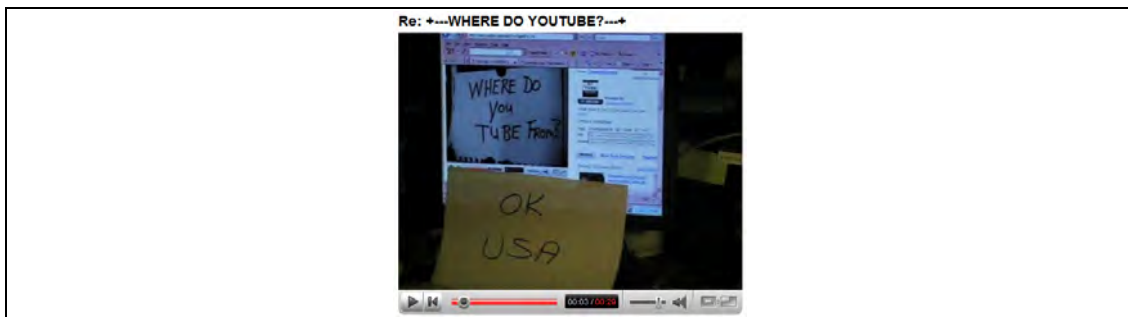


Fig. 36. Response function represented by post it attached to the initial video playing on screen.

All these videos express the relation by recapitulating the request and thus stating the function of the text in the exchange as a response. This is similar to (and indicative of) other forms of asynchronous communication, like letters (when they open by referring to the letter they reply to) or email replies (when, at the bottom, the text of the initial email is quoted). In turn, this normally does not happen in real-time interaction, because it is understood that the next turn refers to the preceding one (unless one takes up the topic of an intervention which took place earlier in the interaction and was then followed by others which passed through to another topic).

In representing (by any mode) the first video and the response link, these responses represent the whole exchange (from the response perspective), so that it is recoverable even if the first video has not been viewed or the response-link on the interface is broken (because of the first video removal, for example, or because the respondent changes the video status and links it to another video).

#### *b. Prospective narration of the exchange*

A further type of these ‘exchange narration’ responses is constituted by the ones (51) which re-launch the topic question to viewers at the end of the video, after providing

their own answer. These videos, rather than (or beside) resuming the previous turn in the exchange, take up the request pattern in *ChangeDaChannel* video (he gives his own answer before asking the question) and prompt responses in their turn, so as to extend the thread to a further level, i.e., to open a new exchange (Fig. 37).



Fig. 37. Topic question relaunched to viewers.

This final question is thus both a cohesive tie with the request (lexical repetition of the topic question) and a prompting cohesive tie with a possible sub-response. It is worth noticing that these request-responses are generally not followed by (many) video responses (cf. Section 4); evidently respondents prefer to link their contributions to the original request.

### 3.3.2 Unmarked organization: Location as New and focus

The answer can be represented in an unmarked information structure, either with the location represented as New (with agent and action as Given) or as focus (by representing only the location), at times extending it with a further rheme.

Most responses (603) represent – in various modes – the three elements of the topic question ‘agent + action + location’. These responses organize the location as New information and focus and a recontextualized repetition of the agent as theme and Given. They can do it by means of language (e.g., ‘I (You)Tube from/in...’) or with images (by filming/portraying the (You)Tuber (You)Tubing at her laptop).

Following Halliday, lexical cohesion is a ‘stronger’ cohesive tie than ellipsis, so that when the location is represented together with the agent (and the action of (You)Tubing), the response has more explicit clues of relatedness than when only the location is represented. Furthermore, as discussed in 3.2, the more attuned the representation, the more related the exchange, so that when the (You)Tuber shows a handwritten paper, the video response is more clearly perceived as related with the initial video than when – say – Homer Simpson is represented wandering in a ‘waste

land’ and asking himself ‘what is this? Shelbyville?’<sup>92</sup>.

42 responses just answer the question by mentioning or representing a location, so that the location is New information (if you consider the request video as Given) and focus, as well as theme (starting point of the response). Here the agent and the action of (You)Tubing are cohesively tied by ellipsis with the request video.

The focus, both here and when also the agent and the action are expressed, can be extended by a further rheme (200 responses), so that, for example, the handwritten paper of the location can precede a slideshow of (more or less) famous monuments. Alternatively, the rheme can add evaluations of the initial video or other information/requests, not related to the thread-topic, according to the (You)Tuber’s interests (e.g., the invitation to subscribe, or the anticipation of future video topics). In this latter case, responding to the initial video is taken as the chance for drawing the viewers’ attention to something of interest to the (You)Tuber. Cf. Fig. 38.

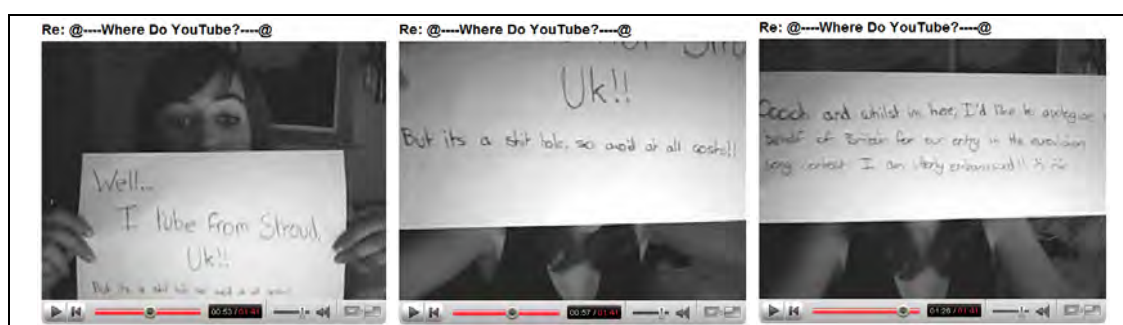


Fig. 38. Topic answer extended by a further rheme.

### 3.3.3 Marked organization: Location as Given and circumstantial element

19 videos *describe* a location or some of its features e.g., in the form of documentary-type videos; in a video a ‘Litchfield views’ typed on screen introduces the filming of landscapes. In these videos the location is theme and Given information and the details on the location are New.

Other 27 videos represent various entities/events with a location mentioned at a certain point, without it being the focal topic, but just a circumstantial element, as in the aforesaid video excerpt of *The Simpsons*, where Homer Simpson walks around and, after awhile, asks: ‘what place is this, is this Shelbyville?’. In a video the (You)Tuber shows how people make coffee in Austria (with ‘Austria’ just mentioned in speech at the end of the quite long filmed demonstration), while its description

<sup>92</sup> Note that in *The Simpsons* TV series, Shelbyville is the antagonist town of Springfield (the Simpsons’ one), considered as worse than the hometown. Significantly, by posting this excerpt as a response, the (You)Tuber communicates the location and, simultaneously, provides a negative evaluation of the town, by intertextually referring to the value of the town in *The Simpsons*.

specifies its intentional topic relatedness ‘at the end of the video i say where i come from :-)’ . Another response is a documentary about ‘Chinese Religion in Taiwan’, as the title says; here the location is circumstantial in the title and Given in the video, since the voiceover starts by saying: ‘Taiwan is ...’ (Fig. 39).



Fig. 39. Marked textual organization.

These two latter types of responses (location as Given and as circumstantial element) are less cohesively tied to the request, not because of the absence of the essential cohesive element (the location), but because they present a textual organization which is marked for an answer to the topic question, i.e., the location is not represented as New information and focus, as the answer would ‘traditionally’ require. As anticipated in Chapter 4 (Section 2.4.4), the marked textual organization is typical (but not exclusive) of responses which were uploaded in a previous time and later linked as responses to the initial video. In this case, videos were not made purposely for the exchange and the view of the initial video (and its achieved popularity) prompted their uploaders to link their videos – which, at a certain point, represented a location – as a response to it.

Although these responses have a marked textual organization for answering the question ‘where do you tube?’, they are nonetheless related to the initial video in that they provide (somewhere in the video) the prompted geographical information as requested. Even more, it is the link of response established with the initial video which foregrounds an element – the location – which would be backgrounded otherwise. In other terms, the recontextualization of the video as a response to the ‘Where Do YouTube?’ one prompts the viewer to perceive the location as salient, even when it is represented as Given or as a circumstantial element.

Furthermore, in the case of the documentary-type and of *The Simpsons* excerpt, they both lack the modality clues of home-made videos (no blurred focus, abrupt cuts, over/under exposition to light, or unstable camera position, etc.). This makes them less ‘personal’ so that the agent and the action (i.e., a (You)Tuber (You)Tubing) are less implied. In generic terms, these responses do not attune with the videoblogging genre of the first video; this, together with the location not in focus but as Given or circumstance, makes them less cohesively tied to the request, so that only the response link on the interface triggers the viewer to search for clues of relatedness.

### 3.3.4 Cohesive ties in the paratext

Clues of relatedness can be also variously represented in the paratext of the video, thus contributing to the degree of relatedness of the response with the initial video (for viewers who read the paratext).

#### a. *The video thumbnail*

The video thumbnail can represent a location (e.g., by means of a map or written on a paper-sheet). The thumbnail can be selected by the (You)Tuber (otherwise, it is automatically given by the interface by using the first shot of a video) and is a cohesive tie of repetition of a snapshot of the video. It generally functions as an eye-catching cue that triggers views to a video.

Although, in selecting the thumbnail, various strategies can be adopted by (You)Tubers according to their interests (e.g., according to what they believe is most salient in the video, to what they think may stimulate the browsers' curiosity etc.), in the response page the thumbnail can be a strong clue of relatedness of a video prior to its view. It is probably the strongest one which can be found in the paratext, since it uses a snapshot of the video and, hence, it is materially related to the video content, differently from the title or the video description. This is particularly true when the thumbnail portrays salient formal elements prompted by the first video, like the handwritten paper in this case, irrespectively of whether what is written is readable or not. Alternatively, when the thumbnail portrays a close-shot face, it is indicative of the videoblogging genre of the video, which attunes with the genre of the initial video (more than a shot of a TV excerpt). For various relatedness clues represented in the thumbnails of the responses, cf. Fig. 40.



Fig. 40. A screenshot of the thumbnails of the responses.

Even if it is the strongest one in the paratext, the thumbnail is only a clue for relatedness, in that it prompts an implicature of relatedness which may be refuted by the video content. Indeed paratextual information can be used by the uploader of a video to mislead the viewer (so as to get more views), by selecting a snapshot which is totally irrelevant to the video topic (and which may have been even inserted in the video for the purpose of using it as a misleading thumbnail)<sup>93</sup>.

*b. The title*

Paratextual clues can be in the title, in the form of ‘RE: Where Do YouTube’ or of an explicit ‘I (You)Tube from [location]’ (or ‘[I (You)Tube from] location’). The first form of the title is a weaker cohesive tie than the latter, since it is the default title which is automatically given when a video response is uploaded to a video. Nevertheless, this default ‘RE:...’ title is a stronger clue for relatedness than a totally topic-unrelated one (e.g., ‘Bad Hair Day – Happy Baby’ in the corpus).

In turn, the reformulation of the topic answer in the title (‘I (You)Tube from + location’) or of the topic question (‘Where I (You)Tube from’) is a stronger clue of relatedness, since it is indicative of the fact that the (You)Tuber has spent some effort in it (i.e., an intentional transformation has taken place). Some videos are titled with the name of a location (e.g., ‘Maybe Shelbyville’, the title of the a.m. excerpt of *The Simpsons*), which functions as topic of the video content (i.e., the typical function of titles). The location in the title helps foregrounding it in the video when the latter does not represent it as focus (as in *The Simpsons* excerpt, where it is a circumstantial element). In the title the location can also be represented as circumstantial element, e.g., in a response titled ‘A Drive Into Boston’, filming the landscape from a car, with no name of the location mentioned in the video (Fig. 41).

*c. The video description*

Eventually, some videos can have an explicit description on the video page, so that they can express (1) the response function of the video in the exchange (‘this video responds to...’), (2) the deictic reformulation of the topic question (‘where I (You)Tube from’), or (3) the name of the location. Note that while the first two types of description need the viewing of the video for the answer to be known (they simply establish a cohesive tie, indicating that the video is related), the latter makes the view of the video unnecessary to know the answer, so that the video description is a selective written transduction of the video.

Video descriptions can also refer to the attuned mode of the thread and this also

---

<sup>93</sup> Indeed, some videos explicitly exploit this device to get views and to playfully tease the viewer by, for example, using a thumbnail whose shapes may be interpreted as sexually related. The view of the video then reveals that the thumbnail’s shot is only a zoom-in onto a larger image sexually unrelated (it could be a close shot of the edges of the thumb and index fingers tight together, for example, which, when zoomed-in, may be interpreted as the image of a bottom or breast).



constitutes a further clue for relatedness; cf. the following description, which highlights also the collective attuning trend in the thread (i.e., ‘Just like the others’):

Just like the others. I won’t say a word. Respect! This is kinda cool, to communicate, isn’t it? It’s ‘unique’!

---

In the thread, the distribution of the elements in the paratext is relatively independent of that of the elements in the text. Thus, videos which have the location represented as Given (e.g., the documentary type), can mention it as New in the description (e.g. ‘[This video is about] where I (You)Tube from’). In these cases the video description functions as a contextualization of the video content. Even more, given that a video usually represents a location, even if not foregrounded (in terms of the setting where the participants and the events in the video are filmed), the paratext explicit reference to a location or to the topic question has the video perceived as topic-related. So, for example, a video featuring a close-shot of the (You)Tuber in her room may be considered as unrelated, but, if its title or description reads as ‘where I (You)Tube from’, it is then understood as the representation of the place of (You)Tubing.

In other words the title – as all titles – has a foregrounding function of what is represented in the video; it guides the viewer to perceive certain elements of the representation as salient. For example, in the case of a video titled ‘Caragua’ filming guys having fun in a swimming pool, it is only the title which gives a clue of relatedness, in that it foregrounds the location where the video is filmed. In an analogous way, a title such as ‘top ten implosions in Las Vegas’ (which presents a location as circumstantial element) foregrounds the location of the video (i.e., a selective mashup of demolitions of skyscrapers), which could otherwise be considered as unrelated (Fig. 41).



Fig. 41. Titles foregrounding the location.

This varied distribution of clues of relatedness in the text and in the paratext evidences to the fact that (You)Tubers use differently their available resources, so as to modulate the relatedness of their response according to their interest. Indeed, if one wants to link a previously made video as a response in the thread (e.g., the above

cited filming of guys having fun in a swimming pool), it is easier to insert cohesive ties by editing its paratext rather than editing the video, i.e., by re-titling it with the location, rather than inserting an overlaid typed ‘Caragua’ in the video.

These instances of video-interaction are exemplary of how the ‘copy-and-paste’ production of semiosis (i.e., selection, transformation and recontextualization of resources according to the sign-maker’s interest), which is so typical of many contemporary forms of communication, reshapes the traditional notions of cohesion and coherence.

*d. The comment section*

A different type of paratextual information can appear in the comment section of the video. Here, the meaning of the video can be negotiated with the uploader, so that some responses have comments questioning their meaning (e.g., ‘WTF?’, ‘what was that?’, ‘I didn’t get it’), with the (You)Tuber replying by expressing the response function of the video in the exchange (‘it’s a response to...’), sometimes adding the link to the initial video page. In other words, clues of relatedness in the comments are given by the (You)Tuber as a response to a prompting comment by a viewer.

### **3.3.5 No (explicit) clues of relatedness**

Eventually, 23 videos represent none of the above discussed clues of relatedness, either in their text or in their paratext. Therefore, they position themselves at the extreme ‘unrelated’ pole of the continuum.

Their uploaders may not be interested in providing a topic-related reply, but, in augmenting the visibility of their video by exploiting an affordance of the medium for linking one’s video to others. Indeed, although posting a response seems not to be a very effective way to achieve visibility (Benevenuto et al., 2008a), the video response link does not prevent any other option available for promoting one’s video (within and outside *YouTube*), hence exploiting it entails very little costs. Even in this case, however, it remains the question of what prompted the uploaders of these apparently ‘unrelated’ responses to link them to that particular initial video, among the many popular ones on the Website.

Furthermore, although, in van Dijk’s terms (1985), these responses construct ‘globally incoherent’ exchanges, *ChangeDaChannel*’s validation signifies that he does not consider them ‘disturbing’ (unlike traditional ‘spam’). Whatever reason drove his approval – including interpersonal ones, e.g., the respondent being a respectable (You)Tuber or a very dear friend of his, so that it would be impolite to deny the response – they indeed build successful exchanges with the initial video.

In sum, something in *ChangeDaChannel*’s initial video prompted the respondents to

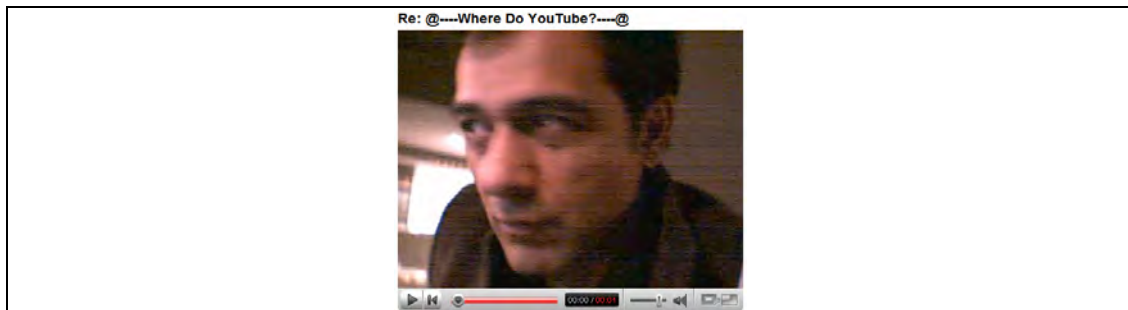


link their video to his, and, in turn, something in these ‘unrelated’ responses prompted *ChangeDaChannel*’s approval, so that this double prompt-response relation, driven by both *ChangeDaChannel*’s and the respondent’s interests (which may well be different), results in incoherent yet successful interactions.

Furthermore, the absence of explicit clues of relatedness does not straightforwardly define a response as ‘unrelated’. Ultimately, relatedness is subjectively perceived, particularly in the absence of explicit or foregrounding signifiers, as illustrated in the next two sections.

*a. Problematic cases*

Some of these responses ‘unrelatedness’ status is indeed problematic. For example, two responses just feature the (You)Tuber’s face for a few seconds (Fig. 42). It is debatable whether these responses were meant to represent where the (You)Tuber (You)Tubes (e.g., the actual spot in the house) or not. No resources are used to foreground the location; neither gestures (for example, a finger pointing towards the wall or a hand showing the environment), nor verbal language (either spoken or written), nor the paratext (e.g., ‘Where I tube’ in the title or in the description), nor a themed soundtrack (e.g., ‘home country home’), not even other signs which may represent the geographical location (e.g., the country flag on the wall or the hat of the town’s baseball team). If these videos were to be considered as actually answering the question, they do it in the most economic way, i.e., they have no redundancy at all in their representation so that no specific signifier of the ‘where’ is made salient.



**Fig. 42.** 1” response featuring the (You)Tuber.

Implicitness is widely practiced in *YouTube* videos (cf. also Chapter 6, Sections 3.2.1.a, 3.3.2.b and 3.4.2.b), so that it is by no means capacious to wonder whether these videos were actually meant to answer the question in a very economic way and leave a great deal of interpretative work to the viewer. Indeed, they employ the minimum resources needed to answer the question – i.e., a shot representing the (You)Tuber (You)Tubing in her room – without framing or foregrounding them.

In another case, the video is titled ‘Common path’ and is made of a slideshow of images portraying various landscapes, with a suggestive soundtrack in the

background. Here again, the video represents locations, so that it is problematic to consider this video as unrelated. However no other clues, either in the video or in its paratext, indicate that the video is meant to answer *ChangeDaChannel*'s question. If one comes across this video through browsing, it would seem a typical 'suggestive' video, portraying emotional images with a sentimental soundtrack and an evocative 'title'. Conversely, if one watches the video through the 'Play all responses' option on *ChangeDaChannel*'s initial video page), the locative relation is more likely to be established. In this case, it is the viewer's search for coherence which drives the interpretation of the exchange, as argued by Halliday:

It is almost impossible to construct a verbal sequence which has no texture at all – but this, in turn, is largely because we insist on interpreting any passage as text if there is the remotest possibility of doing so. (1976: 23)

In all cases, it is reasonable to assume that, in trying to exploit maximally the medium affordances in order to get the video viewed, the respondent chose *ChangeDaChannel*'s video (among all the existing ones), prompted by the common theme 'location' that the two videos share.

Another response features the (You)Tuber speaking fast-forward while making a lot of gestures. Here the speech is not decipherable, so that it is not possible to know whether the question is answered or not. Its video description reads as:

my trying to understand how to be featured in youtube?

The response may or may not be related to the initial video (it may also be intended to show the room where the (You)Tuber (You)Tubes), however the description mentions the purpose of the video, i.e., to be featured on *YouTube* homepage (and so does its title 'how to become a featured video?'). When watching this video, without knowing its response link, one would infer that the content of the fast-forward speech argues on the ways through which a video can be featured. In turn, if one watches the video as a response to *ChangeDaChannel*'s (e.g., through the 'Play all responses' option), one could assume that the speech answers the question and that the 'fast-forward' effect is meant to produce humour and thus differentiate itself from all the others and getting featured<sup>94</sup> (and, maybe, as a means to respond evasively, without actually disclosing the geographical location, as in the cases discussed in 3.1.3). Be as it may, *ChangeDaChannel*'s initial video was indeed featured on the Website homepage, so, this may have prompted the respondent to select *ChangeDaChannel*'s video, among all the existing ones on the Website (i.e., to post a video whose title deals with 'how to be featured' as a response to a featured video). In this case, the response does not relate to the initial video topic; rather it relates to it by taking up a background element (its being featured), making it salient and responding to it.

Another topic-unrelated video takes up the verb of the topic of the initial video (to

---

<sup>94</sup> Cf. the value of being the first in doing something in *YouTube* videos (Chapter 4, Section 2.4.7).

(You)Tube) as a prompt to relate its response (Fig. 43). It features a (You)Tuber holding a cd-rom in his hand, looking at it and paraphrasing the famous Hamlet's dilemma:

youtube or not youtube that is the question

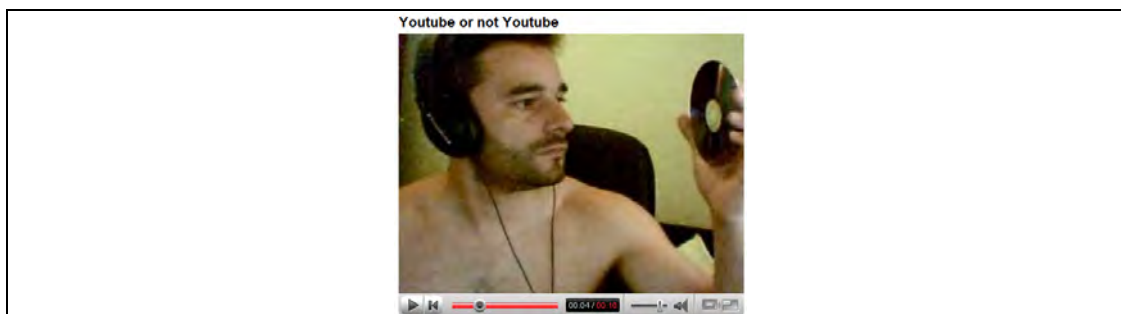


Fig. 43. A secondary element of the initial video as the responded prompt.

These latter videos establish with the initial one an interest-driven prompt-response relation which disregards the topic of the request, takes up a secondary (or background) element of it and makes it the topic of the response. While this phenomenon is not very frequently attested in this thread, so that its description here must be cautious and its instances have been considered among the 'problematic' ones, this strategy is extensively used to establish relatedness in the thread started by the video titled 'Best Video EVER!', as discussed in the next chapter when describing the 'inferential' responses (cf. Chapter 6, Section 3.5).

*b. Interest-driven perception of relatedness*

Before trying to sum up the analyses devoted to video responses, it is worth stressing that, the elements of relatedness between the response and the initial video are to be understood as clues that, when present (and the more of them present), establish degrees of (semantic) cohesion and (formal) attuning. That is to say that they allow the responses to be arranged on a continuum from the ones displaying redundant clues of relatedness to the ones with the least clues of relatedness (in Halliday's terms, which have less cohesive ties). The more the clues, the more a response is likely to be perceived as immediately related to the initial video; even more, in the case of the videos which are placed at the extreme positive pole of relatedness (e.g., the narrative type, even better when also combined with paratextual info), the viewer does not even need to watch the initial video to interpret the exchange.

However, this does not mean that only the videos which have no clues at all will be perceived as unrelated by any viewer. Coherence is subjectively perceived according to the viewer's interest (and of her knowledge of the practices of the specific semiotic space). Therefore, if the viewer is a researcher who wants to see how videos establish relatedness, she will adopt an extensive criterion and will be ready to

consider as related also the videos that are at the ‘least-related’ far end of the continuum, e.g., the videos having a non-geographical location as a circumstantial element mentioned only in the paratext (as in the case of a video with the description reading as ‘chilling out a la oficina’, filming some guys smoking weed and chatting). In other terms, according to her purposes, she will search extensively for any clues everywhere, in the video and in the paratext, so as to avoid filtering out any possible precious data. In turn, a viewer who wants to see how (You)Tubers have answered – e.g., in view of making her own response – will probably consider as related only the videos placed towards the ‘most-related’ end of the continuum, while considering the other ones as an annoying waste of ‘precious’ watching time, like traditional ‘spam’.

As evidence to the subjectivity of the perception of relatedness is the example of a video which lacks all the above discussed clues of cohesion and attuning. The video a close-shot of ‘LOVE STOP The WAR!!’ written on a palm of a hand, without any location represented either in the video or in the paratext. This video might be considered as totally unrelated, and hence incoherent, by many, *but* by the initiator himself. He indeed posted the following comment to the response:

dag i wish I would have seen this before I finished the final:(

In sum, *ChangeDaChannel*’s interest<sup>95</sup> in finding a rhetorical closing has led him to consider the response as related, or, better, as relevant material for his summary.

This example further testifies to the fact that successful communication in video-interaction is driven by the participants’ diversified interests, and that the understanding of the participants’ intended meaning is not at all essential in this practice. In fact, the misunderstanding is actually acknowledged, sought for and practiced; as discussed in the analysis of the video-summary (Section 5), successful communication is conceived as far as it is a productive means of transformation of resources into new texts. Barthes’ visionary *Death of the Author* (1977a) is really foretelling of this new practice.

### 3.3.6 The relatedness-continuum: to sum up

Recapitulating, five elements concur to establish various degrees of relatedness between the responses and the initial video; three of them correspond semantically the topic question of the initial video (i.e., the agent, the action and the location); one signals the interactional function of the video in the exchange (i.e., the response link further represented in the video), while the fifth is constituted by the attuning of the response with the multimodal deployment of the initial video and, more specifically, with the (marked) mode used to represent the topic (i.e., handwriting on paper).

Of all these elements, only the representation of what is needed to answer the topic

---

<sup>95</sup> A frequent reference to ‘peace’ characterizes his persona (cf. his channel slogan: ‘Peace be split!’).

question (i.e., a location, in this case) is essential to establish semantic relatedness. Nevertheless, the marked mode establishes relatedness also in absence of the essential semantic element. Hence responses can either cohere in topic or attune formally, as represented in Fig. 44 (square brackets indicate the optional elements):

<b>1. Topic-relatedness:</b>	[I] + [(You)Tube] + location + [response function] + [mode]
<b>2. Formal attuning:</b>	[I] + [(You)Tube] + [location] + [response function] + mode

**Fig. 44. Clues of semantic (1.) or formal (2) relatedness in the thread.**

On the basis of how these elements are (1) present and (2) textually organized in the responses, various degrees of relatedness can be perceived. So that, on the most-related end of the continuum are the responses which resume the interactional exchange retrospectively (which can be interpreted without even viewing the initial video) or prospectively (which re-launch the topic question to viewers).

In answering, responses can represent all three semantic elements (agent, action and location), so that the location is New and focus and the textual organization is the unmarked one for answering the topic question. Analogously, responses can represent only the location; in this case the location is the focus of the answer, while the other elements are cohesively tied by ellipsis (so that, the initial video must be viewed to retrieve the elliptic elements – the ‘who’ and ‘what’ of the location).

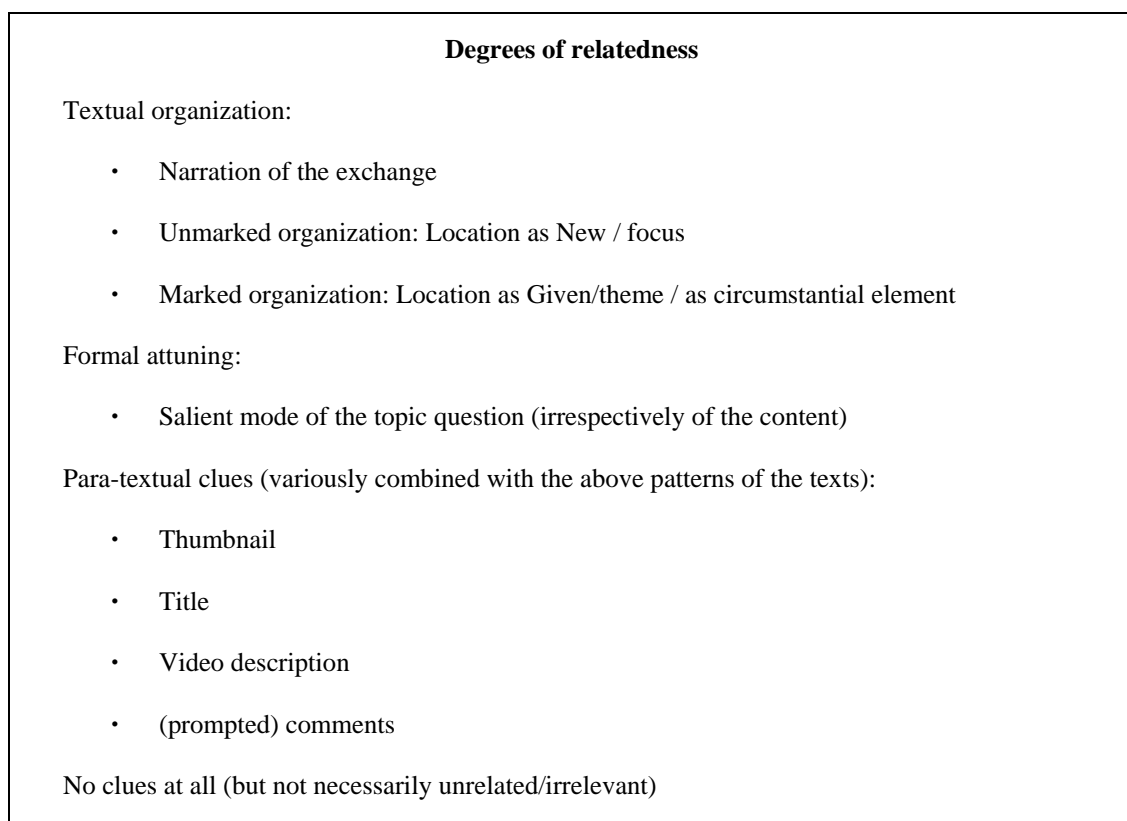
Less cohesively tied are the responses which provide the description of a place, so that the location is the theme and Given, while focus/New is the description (i.e., ‘Taiwan has many religious practices...’). The location can also be represented as a circumstantial element, e.g., represented at some point in the video, while its content focuses on other entities/events. Generally, in these cases, the markedness of the textual organization is the fruit of a response-link established to a video which was created for other purposes than to interact with the request and was later linked to it.

These cases are exemplary of the low cohesion and coherence which characterize the texts of contemporary forms of communication produced through a ‘copy-and-paste’ technique, i.e., through the selection of existing material and its recontextualization in new interactions (cf., also Kress and Adami, 2009).

The representations in the texts can combine differently with clues of relatedness in the paratext. Thus the topic mode (i.e., handwritten paper) can be represented in the video thumbnail, which triggers relatedness even before watching the response. The title can reformulate the topic-question or mention the location; in this way it contributes to signal relatedness and to foreground the location even if the video content does not focus on it. Analogously, the video description can refer to the topic question or can mention the location. Finally, clues can be provided in the comment section, prompted by comments who question the relatedness of the exchange.

In absence of a represented location (geographical, fictional or private), attuning with the marked mode of the topic question (handwriting on paper) functions as a clue of relatedness, which triggers the viewer to interpret the response as related to the initial video. In this type of interactions, form is highly significant (i.e., meaningful).

Finally, when the responses present none of these clues, neither in the text nor in their paratext, they position themselves at the extreme ‘unrelated’ end of the continuum, although their presence in the thread still indicates that the interaction has been successful. In any case, one may always question the unrelated status of the response and search for any – background – element which may have prompted both the respondent to establish the response link to the initial video and the initiator to validate the response in his thread. Fig. 45 represents the various patterns of relatedness evidenced in the thread.



**Fig. 45. Degrees of relatedness in the responses**

The perception of relatedness is always subjective and is driven by the meaning-maker’s interest, so that, while a response representing a location only in the background can be considered as ‘spam’ by some viewers, a response which does not represent a location at all, nor even attunes in mode, can be considered as ‘relevant’ by other viewers, as in the case of the ‘LOVE STOP The WAR!!’ response, which has indeed been considered relevant by the thread initiator, judged as useful material for a positive message to conclude his summary.

## 4 SUB-RESPONSES

In the thread sub-responses are not numerous (33). Two of them are set to private, so that only invited friends can watch them (again, this grants their uploaders both privacy and the responded (You)Tuber's friendship to access their videos).

The topic of the responses is different from the topic of the initial video. Indeed, while the latter asks a question, the responses answer it. Therefore the videos posted as responses to the video responses can be differently topic-related than the responses to the initial video. The next sections analyse various typologies.

### 4.1 Sub-responses answering the topic-question

Most second-level responses (22) are related to the topic question of the thread; in most cases they answer the topic question (while one comments the topic and re-launches the question to viewers). Very often their responded video re-launch the topic question, so that these sub-responses answer both the thread's initial request and their immediate prompting one (i.e., they cohere with both levels in the thread).

Apart from three answering in speaking (and through filming the location), all these sub-responses are analogous to the basic multimodal form of the first-level ones; they represent the topic answer by means of (hand)writing. Thus they attune with the initial video and also with the video they respond to.

One of these sub-responses explicitly mentions the initiator (in the description: 'A Response to changedachannel... I Tube From...'), without referring explicitly to the video it responds to. It seems thus that the initial video has sufficient scope for its question to be answered and directly addressed by a second level response in the thread. Two responses refer to a prompting video (thus reproducing the 'narrative structure', as in 3.3.1) but do not disambiguate the addressee, so that it could be either the video they respond to or *ChangeDaChannel*'s initial one (although one of these attunes with the spoken mode of its immediate responded video, so that it is more formally attuned with it than with the initial video). In other words, there are no clues as to whether these respondents have watched *ChangeDaChannel*'s initial video or not. Five sub-responses address explicitly the (You)Tuber they respond to, while giving more or less strong clues that they ignore *ChangeDaChannel*'s initial video; very often certain elements of their representations attune with their immediate responded video (this could be the mode, as mentioned above). One video, which responds in speech to a paintbrush-written 'I tube from England', explicitly shows evidence not to have watched *ChangeDaChannel*'s initial video; indeed, here the (You)Tuber mentions his not understanding of the thread topic (although conjecturing a possible initial video):

personally I youtube from England but I have no idea what this youtube is for it

was obviously a response to another youtube but still I youtube from England thereyougo!

Finally, the sub-response which comments on and relaunched the question refers clearly to its responded (You)Tuber as to the one who has initiated the thread; indeed, after saying that it is his first video, the featured (You)Tuber talks about the interesting issue of where people (You)Tube and suggests:

to find out more answers go to 'footykid123' he's made a very good video quote r  
e at dash dash dash dash Where Do YouTube dash dash dash dash at very good

Note that he *spells* the 'RE @----' elements of the title (i.e., 'ar, i, at, dash, dash, dash, dash') as if he had interpreted it as the title given by the (You)Tuber, rather than the default one automatically given to video responses. This, added to the fact that he admits to be new at uploading videos, may be a clue that he is not aware that the video he is responding to is a response in its turn. He eventually asks his viewers where they (You)Tube, without giving his own answer; however, he speaks with a British accent, while the responded video writes on paper the location 'England'. The geographical proximity further reinforces the idea that the sub-respondent has come across the video of his neighbour<sup>96</sup> (You)Tuber which has prompted him to respond.

Moreover, these two latter cases evidence once more to the locally-oriented distribution of video-interaction, as found in Benevenuto et al. (2008a).

#### 4.2 Self-responses: topic specification, development and diversion

Eight second-level responses are self-responses (i.e., posted by the same (You)Tuber of the first-level response). One of these specifically refers to the video it responds to by means of an introductory typewriting on screen:

I said "I tube from beautiful nothern Italy"...

This way, the self-response resumes and quotes its previous video (which filmed a piece of paper with a handwritten: 'I Tube from beautiful northern Italy UDINE!'). Then the self-response announces the function of this further response (i.e., to provide further details on the location), by means of the 'exactly' expressed in the typed 'Let me show you where do I exactly tube from...', followed by a slideshow of photos with captions of the name of the places. In other words, this self-response functions as a specification of the topic of the previous one.

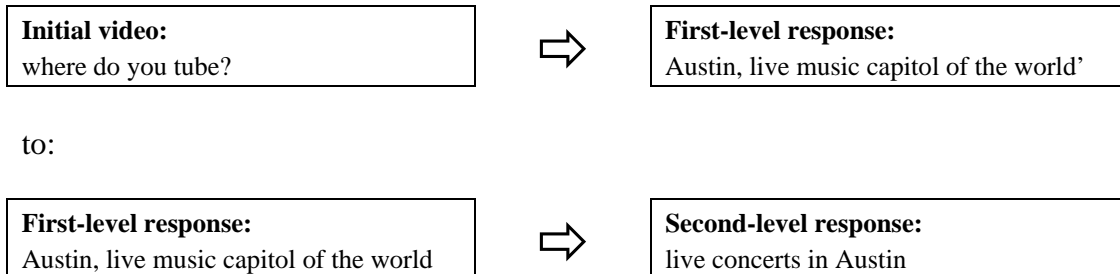
Six sub-responses are self-responses to the same video. Their responded video declares to (You)Tube 'from AUSTIN TEXAS SELF PROCLAIMED LIVE MUSIC

---

<sup>96</sup> National versions of *YouTube* have been launched to which the browser automatically redirects on the basis of the IP location; so that locally-oriented videos are more readily available than international ones.



CAPITOL OF THE WORLD', which is drawn/written on paper, playing with homophones ( 'I' is represented by a drawn eye and 'tube' by the drawing of a tube of 'TOE NAIL FUNGUS 600'). All 6 self-responses feature live concerts in Austin. So these videos extend the thread-topic from:



In other words, the sub-responses take the rheme of the first-level response (i.e., 'live music capitol'; a specification/qualification of the focus 'Austin') and make it the theme of the sub-responses by ellipsis. A topic question is answered, the answer is specified, then the specification is evidenced (indeed the live concerts can function as evidence of the title of music capitol self-proclaimed for Austin). This is a typical instance of topic development in monologic texts (as in written argumentation).

Finally, another self-response (to a video writing 'I tube from uppsala, sweden') films a typed invitation 'ASK TO YOURFRIENDS / TO SAY HELLO TO YOUTUBE!' with a (You)Tuber in a restaurant saying 'Trieste the day before easter welcome to youtube eheh' and filming the guests, inviting them to say hello to *YouTube*. Here the type of relation is not clear; in fact its themed location ('Trieste') is different from the one of the responded video ('uppsala, sweden'); thus it seems that the sub-response deviates from the thread topic, by taking up the focal element 'location' of the first-level response, making it the theme (but changing the reference from Uppsala to Trieste) and developing a completely different topic (i.e., greetings to *YouTube*). Linking his videos together as responses may be a strategy used by the (You)Tuber to get his whole video production viewed.

### 4.3 No clues of relatedness

One sub-response is totally unrelated to the thread topic and responds to an equally unrelated first-level response; the two videos are related between them, in that they both belong to the same genre of 'pathos slideshows' (i.e., photos of affectively connoted entities or events, accompanied by a highly suggestive soundtrack and by an emotional title and description). Two second-level responses give no clue to be related either to the initial video or to their responded one. Finally, the one third-level response is topic-related to its responded video, which is an unrelated second-level response (they both deal with 'alligators' in the modality of reality-documentary).

---

To sum up, sub-responses can answer the topic question either referring to the initial video or to their immediate responded one (showing more or less clues that the initial video is known). Attuning in mode is generally maintained also in these responses, mainly with formal elements of their responded video. Alternatively, second-level responses can comment the topic question or be self-responses, which can further specify the answer given in the first video, or develop and divert the topic. Finally, also these second-level responses can show no clues of relatedness with either the initial video or their responded one. The topic can change as the levels go further in the thread; this evidences to the fact that some (You)Tubers make responses without following the links from the responded video back to anything possibly related to it, while others prefer to know the whole ‘story’ before posting a response.

## 5 THE VIDEO-SUMMARY

Uploaded on 3 April 2007 and notified with an addition to the initial video description, the video-summary is a 3’21” selective mashup of the responses. The video-summary selects, transforms and recontextualizes the resources made available by the responses and condenses the video-thread in a coherent narrative pattern, enhanced by means of a unique soundtrack.

The video opens by framing the narration with a title (‘Where do YouTube’; ‘YouTube’ represented with the Website logo), appearing on the screen of a laptop placed on a toilet. The plot is introduced by a conversational pattern of ‘adjacency pairs’ (Sacks et al., 1974), constructed by shots producing the following sequence:

- ‘where do you tube from?’
- ‘Me?’
- ‘yes you :-]’
- ‘I tube from...’

As shown in Fig. 46, the semiotic resources of the repeated sequence are very varied, so that they communicate diversity and plurality of voices. This polyphonic conversational pattern introduces the topic answers, arranged through a telescoping device, which zooms in from the outer space, to the Solar System, up to Planet Earth. *Google Earth* image of a turning Earth follows, interposed by a handwritten ‘Let me show you around!’. Then, four shots of (You)Tubers (dancing differently yet in time with the soundtrack) lead the way to the geographical locations (Fig. 47).



Fig. 46. Introduction and polyphonic adjacency pairs sequence.

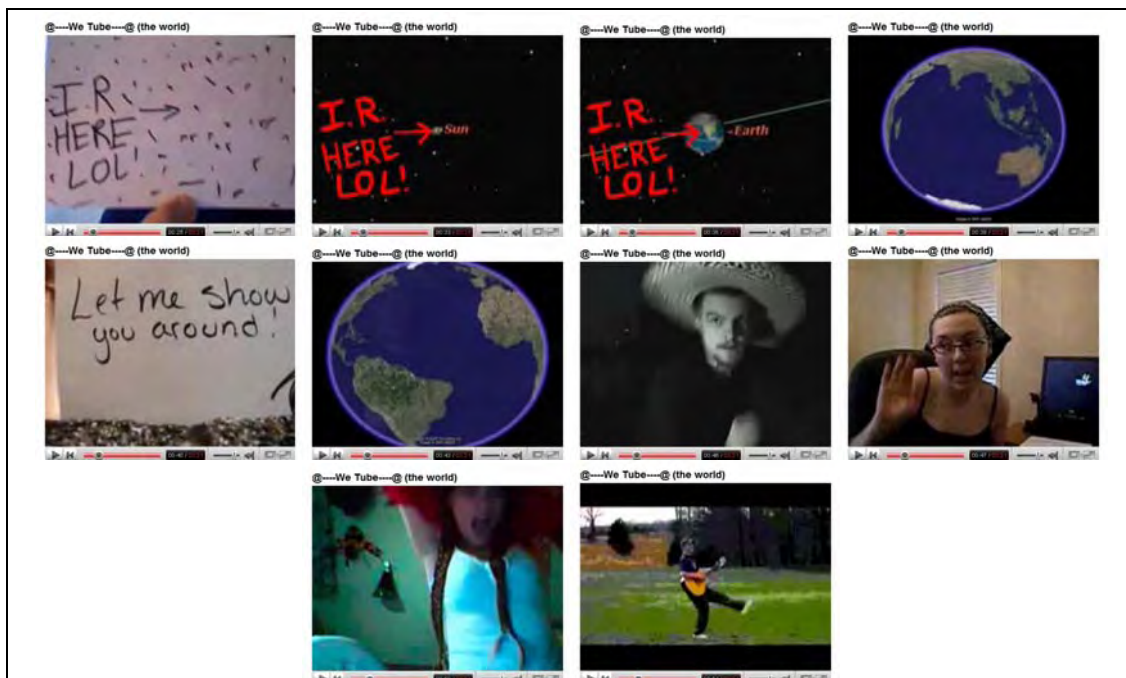


Fig. 47. Telescoping device from outer space to Earth; (You)Tubers dancing.

A random selection of locations functions as a trailer, anticipating a geographically coherent grouping; the Australian shots give way to Europe, then to Asia and the Americas (USA first, then Canada and South America), followed again by Europe.

The geographical cohesion of the locations roundup is further framed by various means; so, for example, a (You)Tuber blowing up a candle marks the passage from US to Canada, which is further introduced by the national flag; the national flag also introduces both the Australian and the US section; a driving car signals the passage from 'Romania' to 'UK', which, in turn, is separated from the Asian section through a '[I (You)Tube] in this little room'.

At the very centre of the video (at 1'39"), in the middle of the U.S. section, the narration reaches its climax: the soundtrack pauses and 'We tube' appears typed on a still image of *ChangeDaChannel* (taken from another *ChangeDaChannel*'s video, not included in the thread); see Fig. 48.

The location roundup concludes with two shots of (You)Tubers taking leave through gesturing 'peace' to the camera, i.e., a V-shaped index and middle finger, the palm of the hand towards the addressee<sup>97</sup>. The soundtrack slows down while a typed 'WeTube' appears on the shot of a guy who slowly turns his back and walks away from the camera (revealing, below a formal grey jacket, his ludicrous boxer-shorts painted with cartoon-like faces).

A black screen precedes the traditional signature tune, made of a (You)Tuber playing his guitar and resuming the thread-topic by singing 'where do you tube?' while the credits scroll from the bottom. In a cinematic finale, 'the end' typed in white fonts on a black screen closes the video (Fig. 49).

---

<sup>97</sup> The peace gesture is a very frequent closing sign of greetings in videos, cf. Adami (2009a).





Fig. 48. Roundup of locations and 'ChangeDaChannel' climax (second snapshot of the fourth row).

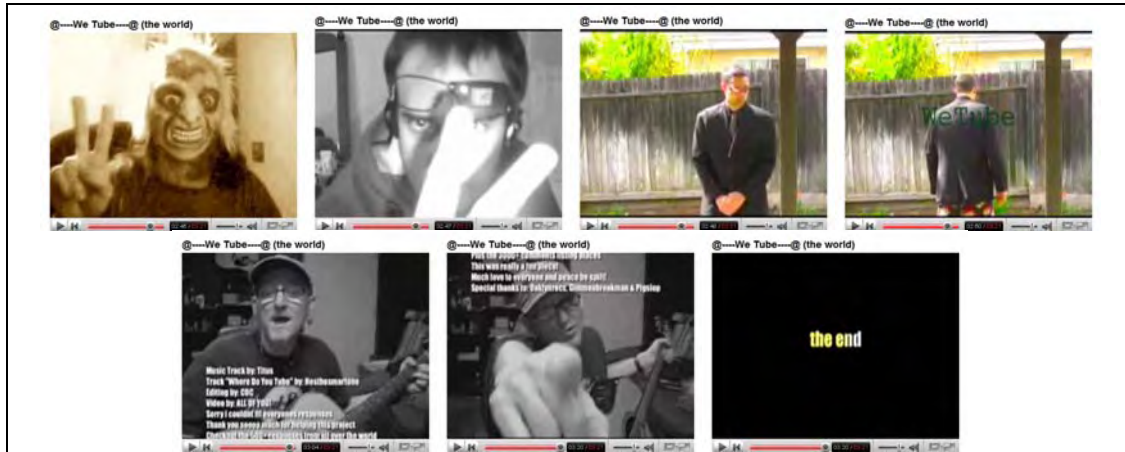


Fig. 49. Peace gestures, guy (in jacket and shorts) taking leave, closing tune and 'the end'.

By selecting, transforming and recontextualizing the representations of the responses, the resuming video presents the video-thread as

- (a) differentiated (You)Tubers;
- (b) coming from all over the world;
- (c) meeting in a space (i.e., the opening *YouTube* logo);
- (d) through a medium (i.e., the laptop);
- (e) gathered around the initiator (i.e., *ChangeDaChannel*'s central position);
- (f) taking part in a polyphonic conversation;
- (g) made by joyful (dancing) participants;
- (h) united in a practice (*YouTube* logo, 'We tube');
- (i) wittily playing with seriousness (the revealed boxers under the formal jacket);
- (j) moved by positive values (i.e., the peace gestures).

Finally, although united, the community is open to anyone willing to join, invited by the final tune re-launching the topic question (this prompt is taken up by some of the responses to the video-summary, which indeed answer the question; cf. Section 6).

The geographical distribution broadly represents the video-thread's (videos representing African countries were posted after the summary was made), and, also here, US locations are more frequently represented. Although a couple of non-geographical locations are shown, the selection seems highly skewed towards the content-specific prompt of the initial video (while 16% of the responses posted before the summary represent a non-geographical location). *A posteriori*, this reinforces the impression that, when posting his initial video, *ChangeDaChannel* was interested in representing the international extent of 'Tuberland'<sup>98</sup> in his summary. He indeed represents his own coherent meaning out of (and transforming) the ones of the responses. In fact, some 'topic-subverting' responses are also included, but they are recontextualized and fit the image of a worldwide *diversified* community.

<sup>98</sup> The term is taken from a response's greeting opening ('Hi tubers from Tuberland').

Consistently with the main multimodal deployment of the thread, the video-summary selects all visual (frequently written) representations, thus further presenting the thread as an attuned and cohesive whole. It also shows a great extent of variation *within* this attuning, perfectly compatible with the coherent whole; not incidentally, the only oral response selected in the summary (the song excerpt), rather than representing a location, resumes the topic and functions as a traditional closing tune.

The scrolling credits define the video as a ‘project’, thank those who took part and helped it, and present the video-summary as a collective work (‘Video by: ALL OF YOU!’). An apology (‘Sorry I couldnt fit everyones responses’) mitigates the effect of the exclusion, simultaneously showing the massive adhesion in the project:

Music Track by: Titus  
Track “Were Do You Tube” by: Hesthesmartone  
Editing by: CDC  
Video by: ALL OF YOU!  
Sorry I couldnt fit everyones responses  
Thank you soooo much for helping this project  
Checkout the 500+ responses from all over the world  
Plus the 3000+ comments listing places  
This was really a fun piece!  
Much love to everyone and peace be split!  
Special thanks to: Oaklynrecs, Gimmeabreakman, Pigslop

The video description rephrases the credits, adding a further interpretation of the summary, which ‘shows the diversity on the one thing that brings us here, YouTube’:

Second Part to "Where Do YouTube From"  
With over 550 video responses from around the world and 3000 comments, this was some work...but fun. This shows the diversity on the one thing that brings us here, YouTube. Sorry I couldnt get everyone in. There was so many! Thanks alot for helping with this project. I hope to do some more stuff like this in the future:)  
Tracks by: Titus & Hesthesmartone  
Editing by: Changedachannel  
Video by: YOU:)  
Special thanks to everyone who took part & also Oaklynrecs, Gimmeabreakman, Pigslop

Apart from the soundtrack and the central image of *ChangeDaChannel*, no further material has been added, so that the summary presented as a multi-authored work of a community representing itself. Yet the editor’s role is crucial; the authorial intervention selects and transforms the semiotic material according to an interest and, by so doing, it makes its own meaning out of the prompted ones.

This interest-driven transformation of the video-thread is totally acknowledged by the participants in the interaction. Indeed, the 207 text comments posted to the video-summary praise it enthusiastically and congratulate with the editor; they may also occasionally either rejoice at their response’s inclusion or lament their – or their country’s – exclusion, in this latter case, always hedging their complaint (e.g., ‘lol’

in: ‘i did a response and im not in it! i dont feel special lol’).

In this genre, conventions have it that the editor’s creative and transformative skills – rather than ‘objective’ summarizing skills – are positively valued. Indeed, many comments interpret the meaning of the video-summary (e.g., by mentioning ‘diversity’, ‘unity’, ‘worldwide’, ‘community’), while no comment ever complains about a response being misinterpreted or recontextualized. This is exactly what a video-summary (unlike, for example, a forum discussion summary) is meant to do, i.e., to make a new meaning out of the prompts given by the responses.

## 6 THE RESPONSES TO THE SUMMARY

Eight videos are posted as responses to the video-summary. The summary does not ask for but rather represents where (You)Tubers (You)Tube; yet four responses still answer the topic question of the initial video. One of them makes no reference to the initial video (and can be meant to take up the question re-launched by the closing tune at the end of the summary). In turn, two responses acknowledge in the video description their anomalous status, i.e., the fact that they answer the topic question of the initial video but are posted as responses to the summary:

Description 1: ‘For Changedachannel! Late entry! Congratulations man!’

Description 2: ‘Hey Changedachannel! Great job! Here is a video I should have done a LONG time ago!’

A further response description states that it functions as a response to the initial one:

I tube from North Platte, Nebraska. Response to "Where do you tube?" Just like to add you really don't need to rate this video. Weirdos.

These responses express their authors’ knowledge of the whole thread and position themselves as ‘late entries’. The summary is not only more recent *chronologically* than the initial one, and thus more likely to be viewed (cf. Chapter 4, Section 2.4.7); it is also the *logical* conclusion of the thread, so that further responses to the initial one would logically remain unwatched. In following both the chronological and logical development of the thread, these responses simulate synchronous interaction.

One response evaluates the summary, by means of the following typed writing overlaid on the shot of the (You)Tuber (who lip-synchs with the writing):

RE: @----We Tube----@ (The world)

Dude, That Was Cool. Cool..

Genius, I'll Say...

Good Idea, I Was In It Too<sup>99</sup>.

---

<sup>99</sup> This remark signals that the respondent had posted a response also to the initial video, which was eventually included in the summary.



All The Hard Work...  
..I Think It Will Pay Off.  
Yes / Cya! / Bye!

A response comments the summary. Titled 'I GET IT!!!!', it features the (You)Tuber speaking 'oh my god i get it youtube is about everyone'. This way, the response gives its own interpretation of the summary topic (i.e., 'youtube is about everyone').

A similar interpretation of the topic (something like 'the community') relates another response to the summary. It films a black and white close-shot of a face appearing and disappearing (by fading the shot into a black screen), while mentioning some ordinary entities ('people', 'love', 'food' etc.), accompanied by a suggestive soundtrack. Towards the end of the video, the (You)Tuber says 'so what is it?', then he fades; he reappears and says: 'youtube?', then fades, then reappears: 'Life?'. Fades. Reappears: 'Both'. Then a typed 'Share the world' precedes the image of the *YouTube* logo. This response is related to the summary by a positive discourse of *YouTube* associated with the world (cf., the summary title 'We Tube(The World)'). In other words, this response disregards the topic of the thread resumed by the summary, while it takes up an element of the summary (a rhetorical discourse on *YouTube*) and makes it salient by responding to it.

Finally, a response is apparently unrelated (it is an animation), but the description mentions the existence of a relation between its content and *ChangeDaChannel*:

WareHouse America iz cool! ~by Mz Hottie \*Check out ChangeDaChannel's  
channel from Warehouse America at:  
<http://www.youtube.com/profile?user=ChangeDaChannel>

Note that, without this description, the relatedness of the response could be perceived only by the ones who share this specific knowledge (i.e., *ChangeDaChannel*'s history and profile, the animation and the relation between the two).

In sum, the initial video request is so 'powerful' to extend its scope also to many responses to the summary (which, in fact, re-launches the topic question in the closing tune). Responses can also evaluate or comment the summary and can even relate to it through an interpretation of its topic which is totally independent of the one of the thread, which generated it. In this case, responses take up an interested prompt of the video summary and respond to it disregarding the thread-topic.

## 7 CONCLUSIONS

The analysis of the 'Where Do YouTube' video-thread has singled out some preliminary features of video-interaction. The initial video sets the ground, i.e., a range of prompts, both in content and form, which can be taken up by the responses.

The thread topic (which, geographically, mirrors an international yet locally-oriented participation) is taken by the responses as a chance for a creative performance, so that the provision of information is paired with – and sometimes overruled by – a playful interpretation of the interaction. The articulated relation between the prompts of the initial video and the responses is a clue of the sign-makers' interest in differentiating their contribution within a given set of possibilities. The exploitation of the topic semantic ambiguity is widely practiced and accepted, so that traditional notions of coherence and relevance seem not to play a great role here, in favour of an interest-driven exploitation of representational possibilities within a given pattern.

The mode, prompted by the initial video and selected by most responses, functions as a tuning device in this highly multimodal interaction, with each response giving its attuned contribution within a generally understood challenge for differentiation.

On the basis of how responses organize textually their content, they can be arranged on a continuum of relatedness, from those which resume the interactional exchange to those which do not deploy any clues of relatedness at all. Among these extreme poles, responses can answer the question with a more or less marked textual organization (including cases which represent the location only as a circumstantial element, while focussing on some other entities/actions).

This differentiated textual organization (and the consequent varied relatedness between the responses and the initial video) testifies once more to the fact that video-interaction follows quite different principles rather than traditional coherence, relevance or cooperation. Indeed, it is the respondent's interest which drives her to post a video (which – say – only marginally mentions a location) as a response to a request which focuses on the location. In turn, it is the initiator's interest which drives him to accept the response thus signalling it as a successful exchange, even when it is totally unrelated with the initial video.

Analogously, viewers driven by different interests (e.g., taking inspiration for their own response or, in turn, researching relatedness patterns, as in the case of the writer) will interpret the clues differently, and thus will narrow down or enlarge the relatedness continuum according to their interested interpretation of the exchanges.

Further levels in the thread can still refer to and answer the initial video or the video they respond to; they can further specify the topic of their responded video; they can develop it or even divert from it, especially by taking up a background element of the prompting video and make it the topic of their response.

With its remixed narrative structure, the video-summary selects the responses and recontextualizes and transforms their resources according to the editor's interest, making a new meaning out of those of the responses. This transformative practice is positively valued by the participants. To stretch Barthes (1977a) a little, here authors *welcome* their death, for their texts to become misquoted resources for new ones.

Finally, the responses to the summary can still refer to and answer the initial video request or can evaluate the summary or, even, can give their own interpretation to it, again, by taking up an element of the video and make it salient in the response, thus totally disregarding the thread topic.

At a theoretical and methodological level, the analysis has evidenced that, in this early stage of its development, video-interaction works on the participants' interest-driven exploitation of the prompts offered by the initial video. The heuristic notion of an interest-driven prompt-response relation proves adequate to account for sign-making patterns in video-interaction.

Throughout the analysis, the notion of cohesion has been used to identify representational elements which signal the relatedness in the thread. It proves to be useful when recontextualized multimodally, so that a cohesive tie can be established between two elements represented in different modes, e.g., the verbal 'Where' can be answered by a drawn map. The analysis has however evidenced to the fact that relatedness can be established also in lack of semantic (i.e., cohesive) elements, crucially, when the response attunes with the salient mode of the initial video. Form is highly significant in video-interaction, which is often understood as a performance, as the (You)Tuber's challenging engagement with the medium.

Furthermore, even when no cohesive ties nor attuning devices are present, video-interaction can be successful, if it fulfils the participants' diversified interests (e.g., as in the case of the 'LOVE STOP The WAR!!' response, considered as relevant by the initiator, who needs a positive message to conclude his summary).

This all again does not totally dismiss traditional notions of cohesion, coherence, or relevance; indeed inferential processes are always at work and coherence is always searched for when interpreting an exchange (the researcher herself searched for the coherence and relevance in the thread and activated a huge bulk of implicatures to organize the responses along a relatedness-continuum). The analysis has evidenced that these notions are useful to discriminate between more or less 'traditionally' related exchanges, yet they are by no means essential for successful interactions (which is not a secondary issue, since these notions were originally devised so as to determine successful communication, cf. Chapter 2, Sections 1 and 2).

Indeed, the ongoing practices of this new type of interaction follow rather different principles. Here cooperation is conceived as an interested participation in the interaction (no matter how coherently organized) and text creation is driven by an interested selection and transformation of resources, both in the case of the summary and in the case of previously made videos reposted as responses to a new one, so that a newly shaped prompt-response relation is established.

In this view, video-interaction is an exemplary form of contemporary communication, where semiosis is produced through a 'copy-and paste' technique

(rather than through transduction or production from scratch). When text creation is made through selection, transformation and recontextualization of resources, traditional notions of coherence, cohesion and marked/unmarked textual organizations are no longer the criteria which discriminate between acceptable and unacceptable interactions.

This deviance from traditional communicative principles, identified in the ‘Where Do YouTube?’ thread only in a minority of exchanges, is rather the ‘norm’ in the other thread of the corpus, which is subject to analysis in the next chapter.

## CHAPTER 6

### ANALYSIS 3/3:

### ‘BEST VIDEO EVER!’ THREAD

*‘The pleasure of the text is our law’*

R. Barthes, *Poe Studies* (1977)

The present chapter analyses the thread composed of the 613 responses (130 sub-responses and 11 sub-sub responses) posted to the video titled ‘Best video EVER!’, uploaded by the *YouTube* account *itschriscrocker* on 17 March 2007.

The analysis of the thread is organized as follows:

- firstly, it introduces and discusses the thread initiator, *Chris Crocker*, since *Chris Crocker*’s character, persona and (You)Tubing history are highly relevant to the type of thread which is built around the initial video;
- secondly, it describes the initial video while trying to point out the possible range of prompts it sets;
- thirdly, it examines the responses, by attempting a taxonomy on the basis of the type of relatedness which they establish with the initial video; for each typology, the various prompt-response relations are analysed (i.e., the different prompts which the responses take up from the initial video);
- fourthly, it introduces and analyses the sub-responses, so as to see the way the thread develops into its sub-levels;
- finally, a concluding section summarizes the results and tries to give account of the different types of relatedness in the thread.

As in the previous chapter, the notion of cohesion is used here as a descriptive tool. However, it is mentioned more sporadically, since cohesive ties are much looser in this thread and often, even when present, do not concur to build coherent exchanges; significantly, here, unlike in the previous thread, even the notion of cohesion proves not to be useful to distinguish between related and unrelated responses. The theoretical consequences of this fact are discussed in the concluding section.

## 1 THE *CHRIS CROCKER* PHENOMENON

Like the ‘Where Do YouTube?’ initial video, the ‘Best Video EVER!’ features its uploader as its main protagonist. Named *Chris Crocker* (username *itschriscrocker*), the thread initiator is a highly discussed *YouTube* celebrity, whose character’s

aesthetics is focused around homo- and/or trans-sexuality and recalls the style of drag queens. *Chris Crocker*'s character has platinum-blond bleached hair, wears make up, (cross) dresses with low-necked and skinny clothes and often dances and moves in front of the camera communicating sensuality, by means of sexy postures and facial expressions. *Chris Crocker* videoblogs often on gay-related issues and on sex-related ones, usually with a dramatic and sentential rhetoric and using very explicit language; this all contributes to make CC's<sup>100</sup> persona quite controversial.

*Chris Crocker* has become widely known after posting a video titled 'Leave Britney Alone' (uploaded on 10 September 2007), in which CC featured (apparently<sup>101</sup>) crying and pleading for the cease of malicious gossip around the singer Britney Spears, of whom CC declares to be 'the greatest' fan. The video soon became viral and sprang outside *YouTube*, widely reported on printed, online and broadcast news. Since then, *Chris Crocker* has started to appear as guest in TV shows, filmed at movie premiers and even at red carpet-situations where CC's idol Britney Spears was present. Eventually *Chris Crocker* has currently been asked to lead a TV show<sup>102</sup>.

### 1.1 The semiotic activity around the initiator

Since CC's outburst of popularity, the activity on the Website around this *YouTube* celebrity has become wide and diversified. For example, a further *YouTube* channel was created on 24 February 2008, named *exclusivecrocker*, (allegedly) based in UK (while *Chris Crocker* is based in US), which promises to show

THE VIDEO'S CHRIS DOESNT WANT YOU SEEING!

Welcome to Exclusive Chris Crocker. You have reached the utlimate channel for all exclusive videos associated with "edutainer" Chris Crocker.

Subscribe to be continually updated with those videos Chris regrets posting and/or the videos that represent Chris in a more realist light. Furthermore the majority of videos that are placed on this channel are "exclusive" videos of the "Crocker" in what is seemingly unfortunate situations.

(<http://www.youtube.com/user/exclusivecrocker> Retrieved 11 November 2008)

Although the channel presents itself as a 'paparazzi-style' one, in which unauthorized material is posted (cf. 'THE VIDEO'S CHRIS DOESNT WANT YOU SEEING!'), most of its posted videos feature *Chris Crocker* videoblogging indoor. Therefore, it is more than plausible that, far from being an unauthorized channel, *exclusivecrocker* is administered by the very same *Chris Crocker*, who provides the content – which, representing CC talking to the camera, could not be accessed otherwise; at worse, it is certain that *Chris Crocker* waives on the existence of the channel.

<sup>100</sup> Given the declared gender ambiguity of the character, no gender-specific pronoun is used here to refer to Chris Crocker; the acronym CC is used instead.

<sup>101</sup> Whether CC's tears were real or not is a widely debated issue among (You)Tubers.

<sup>102</sup> On the porosity of *YouTube* to the corporate media business, cf. Burgess and Green (2008).

The channel provides a description of *Chris Crocker*:

In brief, Chris Crocker is an openly gay American Internet personality and self-described "edutainer" who produces and acts in transgressive videos.

The Tennessee-based Crocker, a stage name, keeps his identity and exact location private because according to him there are safety concerns and death threats in response to his YouTube and MySpace vlogs and profile. According to his MySpace homepage he is based in Los Angeles, California as of January 2008.

Crocker's criticisms are also reported:

Crocker's detractors and critics have accused him of melodramatics, histrionics, and using Spears' personal shortcomings to bolster his own fame. Others have accused Crocker of acting in the "Leave Britney Alone" video, although he insisted that it was a genuine "blog from the heart" on a September 20, 2007 appearance on the Maury Povich show.

Psychologist Kevin Leman noted in an interview that pointed to Crocker, that voyeuristic fascination with celebrity "says that we live in a morally corrupt society" and that younger people seeking fame, like Crocker, are hedonistic attention-seeking mongers

*exclusivetrocker* is not the only channel on *YouTube* which is related to the personage, and other channels exist which (re)post further CC's material, and either homage or detract CC, e.g., *ChrisCrockerResponds* (which posts further CC's videos), *chriscrockermyspace* (which uploads CC's MySpace videos), *itsmschriscrocker* (which reposts videos uploaded on CC's Website); *ChrisCrockerDeleted* and *ChrisCrockerDeleted* (two channels devoted to CC's deleted videos). Others present themselves as CC's fan-channels, e.g., *JIGS2W* (whose description congratulates CC for being charted as the most subscribed channel), and *CrockerTV*, self-declared authored by 'Casey Chris-Crocker-Fan No.1', which dedicates its activity to *Chris Crocker*:

This channel is for :  
Chris Crocker !!!  
The superstar of YouTube!  
Actually, he is bigger than YouTube!

Among CC's self-declared detractors are the channels *ChrisCrockerIsWANKER* (besides the channel's name, cf. its description: 'I just really hate Chris Crocker'), *Chriscrockerhater2k8*, *VERSUSchriscrocker* (cf. its description 'My hate 4 chris crocker is imaginable..'), and *gimmemoresucks*, whose long and detailed channel description mentions:

Yes, I am a real Chris Crocker hater. I'm not one of his lame crew pretending to hate him. I'm a true hater.:)

Self-declared 'haters' of Chris Crocker are also *cutepuppie2007* (cf., its description 'We don't like you Chris Crocker, leave us alone'), *likewowwee* (cf. 'Leave my username alone Chris Crocker/Crocker Crew get a life!'), *treehuggersunited* (whose profile picture features a snapshot of the 'Leave Britney Alone' video

inserted in a 'Wanted Dead or Alive' poster), and *leaveusalonecreep*.

Many other CC-related channels post remixes and spoofs of CC's videos which cannot always be clearly distinguished between homage and hostile spoofs<sup>103</sup>. CC's remixes and spoof videos are uploaded by, e.g., *ChrisCrockerRemix*, *chriscrockerjunior*<sup>104</sup>, and *ChrisCrockerReplies* (which uploads more clearly hostile remixes of CC's videos). Hardly categorizable – sometimes in spite of their name (which, judging from the channel's contents may well be ironic) – are also the CC-related channels *ChrisCrockerSupport*, *LibbyCrocker* and *itschriscrockerxx*.

Eventually, in the attempt of exploiting CC's popularity and gain visibility, many channels use 'Chris Crocker' as a keyword in their descriptions, so that, a search of the string 'Chris Crocker' results in 278 channel<sup>105</sup>; among these, *ChrisCrockerIsED* expressly admits this interested exploitation of CC's name in its channel description:

Yes, I'm an anonymous YouTuber that is EDsploiting Chris Crocker

Outside *YouTube*, *Chris Crocker* has a wide online activity, including a *MySpace* profile and a Website (<http://www.mschriscrocker.com>), which hosts a blog, pictures and videos, and sells CC's merchandise. CC has also made a single, titled 'Mind in the Gutter', which is available for purchase on *iTunes*.

In sum, *Chris Crocker*'s semiotic activity and phenomenon inserts itself within the contemporary *Société du Spectacle* (Debord, 1967), which is newly shaped in terms of diversified and controversial – rather than univocally meaningful – practices through different media by means of 'memes' (Blackmore, 1999; Dawkins, 1976) which spread virally<sup>106</sup> within and outside *YouTube*. In this context, visibility – rather than understanding, consent and sympathy – is an aim by itself, which is pursued by fostering an heterogeneous and often conflicting participation to the spread of the meme. Participants contribute to its spread regardless of whether they are supporters or detractors – or even 'exploiters' – of the content of the meme, thus instantiating the famous maxim of *The Picture of Dorian Gray*: 'There is only one

---

<sup>103</sup> In this regard, cf. Willet (forthcoming); cf. also the discussion in 3.3 and in 3.4.

<sup>104</sup> Cf. its channel description:

Hello everyone I am Chris Crocker Junior thats right chris crocker junior! Chris Crocker may have left youtube but his name hasn't! The empires that have crrumbled will be rebuild, and all those people that mourned all these weeks will mourn no more! [...] Well i am chris crocker junior and i came to youtube for one reason to show that the chris crocker name still lives on!

<sup>105</sup> Data retrieved 12 November 2008.

<sup>106</sup> Cf. Burgess (forthcoming):

Similar to the scientific usage in meaning if not analytical precision, in contemporary popular usage an internet 'meme' is a faddish joke or practice (like a humorous way of captioning cat pictures) that becomes widely imitated. In this popular understanding, internet 'memes' do appear to spread and replicate 'virally' — that is, they appear to spread and mutate via distributed networks in ways that the original producers cannot determine and control.



thing in the world worse than being talked about, and that is not being talked about' (Wilde, 1891). Referring to CC's haters, in her *MSNBC* article linked to the 'Best video EVER!' description (and so endorsed by CC), Popkin writes:

"They just don't get me," Chris says, and it's fairly believable when the young man claims he doesn't care. As long as people are looking, that's all that matters.  
(<http://www.msnbc.msn.com/id/23748983/> Retrieved 13 November 2008)

Here again, coherence, cooperation, relevance and mutual understanding (cf. Chapter 2, Sections 1 and 2) are not essential elements for the chain of semiosis to be successful, as far as it achieves its participants' diversified goals and responds to their varied interests.

Bearing in mind this general introduction on the '*Chris Crocker* phenomenon', the next section examines the initial video in detail.

## 2 THE INITIAL VIDEO

Posted on 17 March 2008 by *itschriscrocker*, the initial video, titled 'Best video EVER!', features the very popular (You)Tuber, finely wearing make-up and a green blouse, filmed at a close-shot, holding the camera with the forearms stretched in front of the face, as shown in Fig. 50. The video is very short; it only lasts 4 seconds, while *Chris Crocker* looks and smiles at the camera and blinks twice. No soundtrack is playing and CC does not speak, so one can hear undistinguished environmental noises in the background (the typical ones of silent scenes recordings). The background, a whitish sheet leaned just behind CC's head, is the same used in the video which made CC famous, the 'Leave Britney Alone' video. The colours of the video are saturated, with high contrast effects and enhanced lightning, so that the 'modality' (Kress and van Leeuwen, 1996, 2006) used in the video is the typical one of 'glamour' shootings and glossy magazines' photos<sup>107</sup>.

---

<sup>107</sup> It is worth noticing that – as mentioned in Chapter 5, Section 3.3.3 – the modality of realism of home-made videos has its conventional 'naturalism' in low-definition, non-professional camera angle and lightning, dramatic cuts or no cuts at all. These modality features have become 'generic' (i.e., indicative of a genre) as a consequence of the most-common sign-making practices resulting from the affordances of home-made videos, which, compared to professional videos, are given by poorer tools available, poorer expertise of the authors and poorer design of the videos. The typical modality of videoblogging (cf. Chapter 4, Section 3.1.3) adds to these features a fix camera angle and the protagonist addressing the viewer through the face – rather than through the gaze trajectory – towards the camera (because of the webcam affordances, as discussed in Adami, 2008a). These features are however changing as home-made video is becoming a widespread genre (i.e., more authors who produce more videos and more viewers who watch them online), so that tools are more and more sophisticated and knowledge is shared. As a result, as testified by CC's and CdC's initial videos in the corpus, home-made video modality features are often akin to those of professional videos. In turn, the very same popularity of home-made videos (to which the *YouTube* phenomenon

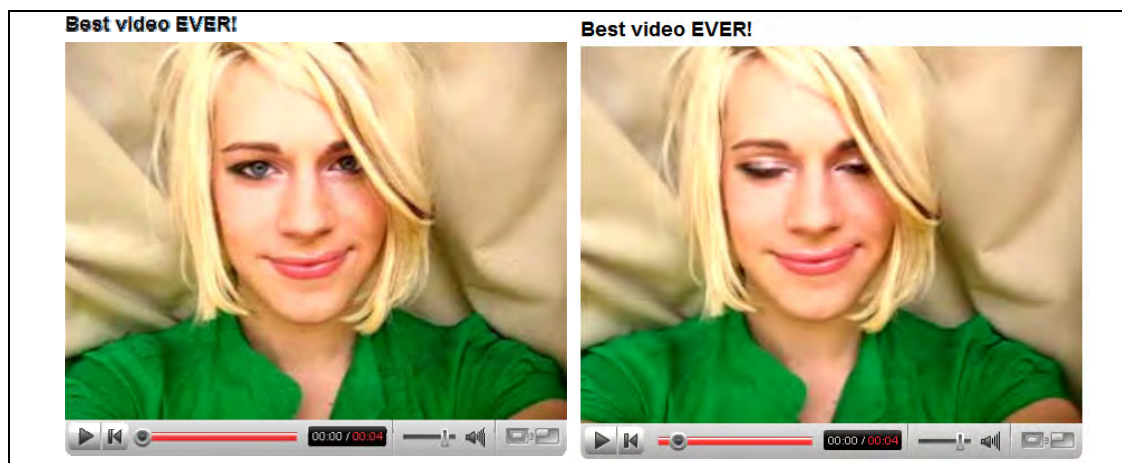


Fig. 50 Snapshots of 'Best Video EVER!'.

The 'Best video EVER!' video description on the homepage reads as: '1 million views for each blink'<sup>108</sup>, which gives some clues on CC's interest in posting the video (i.e., I am so famous that even when I do something totally meaningless I get a huge attention). If one clicks on the 'more info' link in the 'About This Video' section, a further line of description appears, which invites viewers to read the above-cited *MSNBC* article by Helen A.S. Popkin:

Read the new article about me on MSNBC right now here:  
<http://www.msnbc.msn.com/id/23748983/>

Popkin's article celebrates *Chris Crocker* as an 'internationally-recognized icon' and CC's videos as

the next generation of Andy Warhol's cult of personality, in which small gestures become monumental. Unlike Warhol, the 20-year-old Chris belongs to a media-savvy generation that instinctively knows how to work a camera. Like other YouTube celebrities, Chris uses this young medium to become his own media outlet of which he's not the product, but a mirror or a lens.

(<http://www.msnbc.msn.com/id/23748983/> Retrieved 13 November 2008)

On 31 march 2008, the video was charted as the 19<sup>th</sup> Most Responded of all times

---

has greatly contributed) is leading professional videos to adopt some of their features in order to seem more 'realistic' (e.g. in the case of TV ads, documentary-films and reality-shows).

<sup>108</sup> At some point before the data were collected the description must have changed temporarily, since many comments at that time (end March) point to a description saying 'The point of this video is to spread AIDS awareness.', e.g.:

I love how a minute ago, the description said he was trying to prove he could get views by simply blinking, and now he says its for AIDS awareness. :P.

It may have been a view-catcher device, changed after many comments had questioned its relation with the video, e.g.:

spread AIDS awareness...? you look at a camera for 5 seconds and thats spreading AIDS awareness...right. so that means i can take a video of me eating a banana and that will spreading world hunger awareness i guess

(while *Chris Crocker*'s 'Leave Britney Alone' was still the 5<sup>th</sup> most responded video, with 2,110 responses, many of which parodies and remakes of the original video). The data of the thread were retrieved on 4 April 2008, about three weeks after the initial video was uploaded, when it had already gained more than the foretold one million views per blink (2,824,650 views), 31,847 text comments and 613 responses.

Later in May, the title of the video changed in 'Watch Chris Crocker blink' and its description read as:

The point of this video is to show the people who say I am not popular anymore & the people that say I have to be outrageous to get the attention I do- this is to show them that not only am I still popular, but I don't have to do anything other than blink.  
A million views for each blink.

In sum, the new description represents the interpersonal meaning of the video, i.e., *itschriscrocker*'s intention in posting it to prove CC's ongoing notoriousness.

## **2.1 The ideational and interpersonal meaning of the video**

Due to its brevity, the ideational meaning of the video is describable quite straightforwardly, yet its interpersonal meaning and the range of prompts within which the responses may be actualized are less determinable.

First of all, differently from the 'Where Do YouTube?' initial video, the 'Best Video EVER!' is not a request, in that it does not ask for video responses. It also differs because it has no verbal language at all and, maybe more importantly, no topic or theme is expressed through language; indeed the title – which typically summarizes the topic – is by no means related to the ideational content of the video but is rather an evaluation of it ('Best Video EVER!'). Instead, the later changed title ('Watch Chris Crocker Blink') mentions what can be considered the ideational topic of the video (i.e., *Chris Crocker* blinking), while inviting viewers to watch it (i.e., the request is not at video-interacting but rather at watching). However, this revised title appeared after the responses collected for the analysis had already been posted.

If we consider the ideational elements of the video, we can however attempt at interpreting it. Firstly, the 'glamour' modality presents *Chris Crocker* with an aura of 'star-system' membership and alludes at the popularity of the character. Secondly, the widely known 'Leave Britney Alone' background points to CC's most discussed video, which gave CC's fame (or 15-minutes exposure, as CC's 'haters' would say). Thirdly, CC's smile is quite serene and self-confident, which combines with the self-confidence of the type of gaze (and the levity or frivolity of the blinks).

In any case, unlike the 'Where Do YouTube?' initial video, this video is less interpretable if one lacks knowledge of the (You)Tuber's history, past deeds (within

and outside the Website) and discussed status. This means that, in absence of this background knowledge, the video represents a blink as the salient element, i.e., a quite insignificant act. Hence it may just prompt puzzlement (i.e., what does it mean?) and, of course, puzzlement may in turn prompt the viewer to search clues for relevance – in Sperber and Wilson's (1986) terms – in the context (from the video description up to its author's profile and past videos). In other terms, the video's vague interpersonal meaning<sup>109</sup> prompts further information to be searched.

## 2.2 The further meaning of the paratext

By linking the video content to its (former) title and description, a further possible interpretation of the video arises. The video description 'one million views for each blink' is an assessment of CC's self-awareness of popularity on *YouTube*. Indeed the video description does not ask for views or responses; rather, it *states* what in fact is a prediction (i.e., that the video will achieve one million views per each performed blink) in terms of truth value, as of a given fact (no modality markers are there to express prediction, e.g., the modal verb *will*). Therefore both CC's facial expression in the video and CC's prediction asserted as a given fact in the description combine in communicating CC's self-confidence of CC's status as a celebrity.

The title proclaims an exceptional quality ('Best video EVER!') of the video, in which CC does not 'do anything other than blink' though, as per CC's admission in the edited description. Thus the title functions as a view-catcher<sup>110</sup> so as to grant CC the wished 'million views for each blink'; not incidentally, the title has been changed into a more topic-related one (i.e., 'Watch Chris Crocker blink') only after this goal was reached.

Hence one could say that the posting of this video is *Chris Crocker's* assessment of self-awareness of both CC's celebrity status and of the viewing practices on *YouTube*. CC's foretelling is more than fulfilled, and in November 2008 the video has reached 5,561,448 views, so that this video proves that CC's self-confidence is indeed well grounded, as a comment rightly suggests (i.e., 'I bet you still get a million'), after having stressed the view-catcher function of the title:

Well... naturally we're gonna look because of the title. Next time, tell people all you're gonna do is blink and THEN see how many views you get. I bet you'll stilli> get a million =]

---

<sup>109</sup> The ideational meaning is not vague, i.e. CC watching the camera, smiling and blinking; it is the interpersonal meaning of what CC does (the why and for whom) which opens to an undetermined set of possible interpretations. Note that the meaning of the video is vague (i.e., undetermined) rather than ambiguous (i.e., which can mean many different things). While CDC's topic question 'where do you tube' has potential ambiguity in the locative interrogation, which can mean more than one thing (i.e., 'where' as the country, as the exact spot, etc.), CC's blinking is vague since it leads viewers to ask themselves 'why? what does it mean?', rather than 'does it means this or that?'.  
<sup>110</sup> A comment points it out: 'He's called it BEST VIDEO EVER to grab your attention , okay hunz ?'

The later edited description adds a further meaning to the video. Indeed it explains the interpersonal meaning of the video, i.e., CC's intent in posting it (cf. its very beginning: 'the point of this video is...'). In fact, it gives a second reading or a further semiotic space where to add a further meaning to the video.

In semiotic terms, the video and its description are two different and intertwined representations, the latter of which refers to the interpersonal meaning of the former. The description is a meta-representation (i.e., it talks about another representation) which selects and transduces – according to its uploader's interests – certain criterial aspects of the video in written language. In being a transduction, the written description transforms the signs (forms and meanings) of the video and creates a further, different meaning. Therefore, the description is not to be taken as *the* interpersonal meaning of the video, but rather as a further representation which relates to the video and which, together with it, constitutes a complex sign leading to a different interpretation of the video itself. Indeed, CC may have posted this second description as the fruit of a second-thought, after having read the many perplexed comments to the video questioning its meaning (e.g., 'WTF?'); in this case the description is CC's further interpretation of the video.

In any case, the three representations taken together, i.e., the video, the title(s) and the description(s) concur to build a complex sign which signals CC's intention in posting the video, i.e., a stress on the 'phatic' function (Malinowski, 1923) of the semiotic act, rather than on the significance of its content (which is indeed presented as insignificant; cf. 'I don't have to do anything...').

### **2.3 A possible range of interpersonal prompts**

Of course CC's overall conveyed self-confidence of (who does not need to produce any significant content, but whose simple existence suffices for a great deal of communication to take place) can give rise to either rage or admiration and thus can prompt various responses. The relation between the title and the content prompts in itself puzzlement and maybe either irritation or admiration. Cf., in this regard a comment to the video:

HA HA.. I love this. I don't know why but it's killing me. I can't stop laughing. I think when you call something "Best Video Ever" and you just sit there and blink with a smile... it is truly the best video ever... HAHA!!!!!!!!!!!!!!

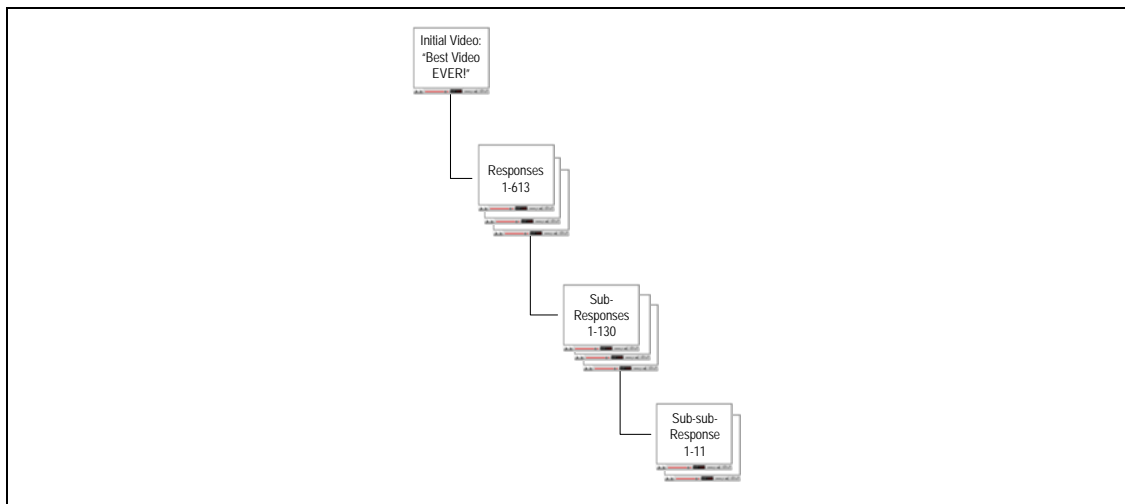
The video can be interpreted as CC's self-confident raising above the viewers, which, of course, as all asymmetrical relationships, prompts either a challenge to or an acceptance of the power divide. Indeed, on the basis of the viewers' diversified acceptance of CC's higher status (i.e., their placing on CC's side, as fans, or against CC, as opponents, or even above CC, as 'outsiders', who raise themselves above the phenomenon and exploit it), the prompt can be interpreted as:

- a challenge addressed to CC's haters (among which the viewer is not); in this case provoking admiration and even support (if the viewer is CC's fan);
- a challenge addressed to themselves, thus provoking rage and conflict in the viewers, so that, by being challenged, I, as a viewer, need to prove that (1) I am better than CC and/or that (2) CC is no one;
- eventually, if the viewer puts herself outside the pro- or against CC's field, the video may just be used as 'material', as a resource to be exploited according to her purposes.

Nevertheless, the absence of an explicit request in the video makes it hard to determine its possible range of prompts and how responses can relate to it. More precisely, not only the absence of a request for responses, but the stress on the phatic function of the video and on the insignificance of any content-related type of interaction determines a particularly wide range of prompts. In other words, when the prompting semiotic act emphasizes its phatic nature, the range of acceptability of the type of response widens. The next section analyses the types of videos which respond to CC's, and how these relate to it and actualize different prompts.

### 3 THE VIDEO RESPONSES

In less than three weeks, CC's 'Best Video EVER!' <sup>111</sup> collected 613 video responses. 44 of them got responses in their turn for a total of 130 sub-responses. Five sub-responses were further responded by 11 videos, thus creating another sub-level in the thread. The thread structure is represented in Fig. 51.



**Fig. 51 'Best Video EVER!' thread structure.**

The amount of sub-responses is proportionally higher here than in the 'Where Do

<sup>111</sup> From 17 March 2008, the day it was posted, to 4 April 2008, the day the data were retrieved.

YouTube?’ thread (cf. Chapter 5, Section 1). This may be explained by the fact that this thread is not started by a request; indeed whereas a video request is more likely to be responded directly, a prompting video can more easily prompt a discussion which can be potentially enlarged to other sub-levels, as mentioned in Chapter 4 (Section 3.1.2) and evidenced in the analysis of the sub-responses (Section 4).

As detailed in Fig. 52, among the 613 responses, eight are set to private, i.e., only invited friends can watch them (this, as observed in Chapter 5, Section 3, may be a strategy used by the respondents to obtain the initiator’s friendship). Seven responses were later removed by the uploader and one was removed due to copyright violation.

<b>Responses</b>	<b>613</b>
Set to private	8
Removed by the user	7
Removed for copyright infringement	1
<b>Viewable responses</b>	<b>597</b>

**Fig. 52 Viewable vs. unwatchable responses in the thread.**

<b>Type of relatedness</b>	<b>No. of videos</b>
Correspondence	125
Commentary	97
Remake	49
Original spoof	34
Inferential	104
Secondary reference to CC	17
Randomness	28
Paratext	27
Unrelated	196
<b>Total<sup>112</sup></b>	<b>677</b>

**Fig. 53 Typology of responses according to their relation to the initial video.**

Set apart these 16 inaccessible responses, the remaining 597 relate in different ways to the initial video. Fig. 53 lists a categorization on the basis of how responses relate

<sup>112</sup> As inferable from the total in Fig. 53 (677, out of the 597 accessible responses), some responses fall into more than one typology. Indeed, as detailed in the related sections, sometimes a ‘corresponding’ response displays also elements of ‘original spoofing’ (this happens, for example, when the (You)Tuber faces the camera and blinks while wearing a blond wig, thus simultaneously ‘corresponding’ the content the initial video and parodying Chris Crocker) or of a ‘commentary’ (e.g., when the blinking (You)Tuber makes also a ‘finger’ gesture, thus insulting Chris Crocker).

to the initial video. Each typology is analysed in the following sections.

Compared to the ‘Where Do YouTube?’ thread, here responses are much more diversified both in contents and forms. This may be due to the absence of a precise topic question in the initial video (and to the vagueness of its interpersonal meaning).

Also differently from the other thread, the ‘Best Video EVER!’ one records a higher number of apparently (cf. Section 3.9) unrelated videos (196). This leads to suspect that the thread initiator is not interested in exercising any censorship on responses, which is consistent with the phatic function of the initial video.

### 3.1 Corresponding responses

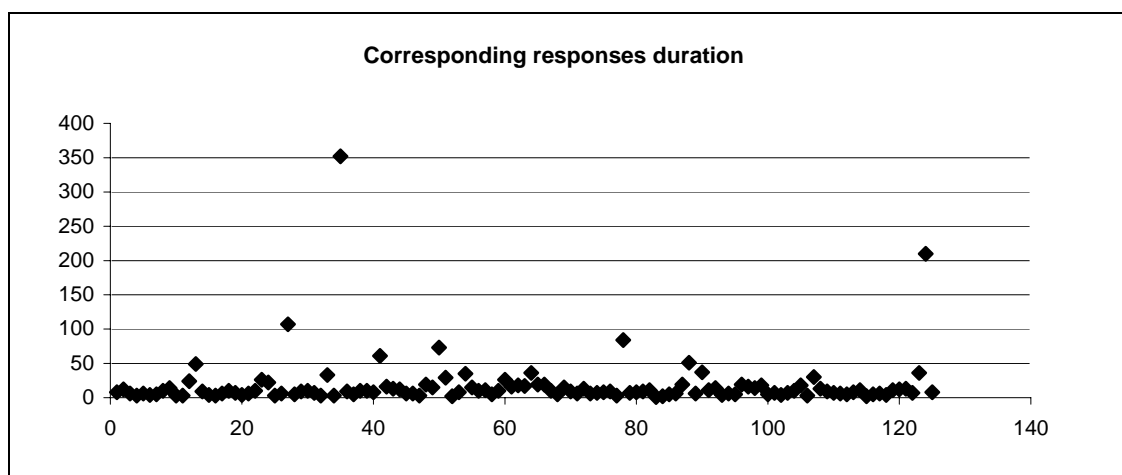


Fig. 54 Distribution of corresponding responses according to their duration: 125 videos; mean 18”.

A large number of videos (125 responses) deploy the same multimodal pattern of the initial one, i.e., very short (cf. Fig. 54) and silent videos with the (You)Tuber facing the camera. As discussed here-after, many of them correspond the blinking of the initial video, while others take up another element of the video (the insignificant action, the sexual connotation, the smile, and so on) and make it salient by corresponding it. Beside the great extent of variation among these responses, they are here labelled ‘corresponding’ response precisely because they deploy the same multimodal pattern of the initial video (they maximally attune with it). Together with it, they build an adjacency pair (as in greetings exchanges, like the wave gesture responding to another one), which is differently modulated on the basis of the element taken up by each sub-type of response.

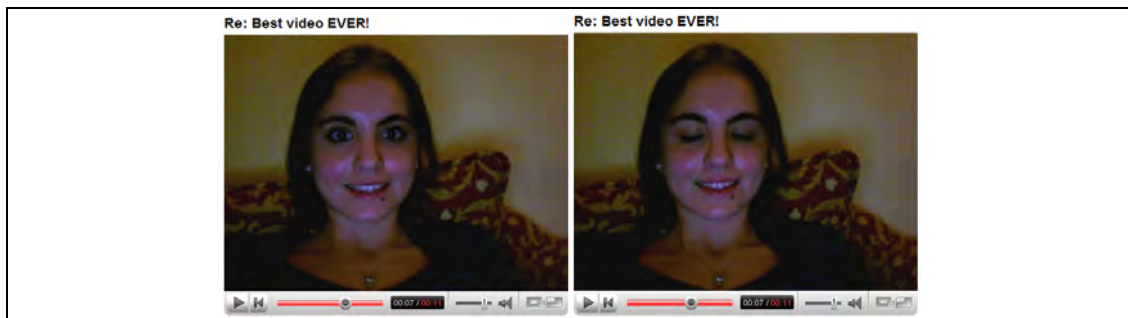
#### 3.1.1 The blink as a corresponded prompt

In 54 responses the (You)Tuber blinks, once, twice or several times; in some cases



the blinking comes at the the beginning of the video, in others at its very end, while, other times the blinking is repeated for the whole duration of the video. Indeed, analogously to the ‘Where Do YouTube?’ thread, even the most attuning type of responses displays a lot of variation in performance.

In most cases, the responses’ interpersonal meaning towards CC is hard to define. Indeed, the responded blinking – in itself a cohesive tie of repetition – could just be an adequate closing turn of an adjacency pair and, in this sense, its interpersonal value would be of cooperation/solidarity with the initiator. In turn, it could also have a parrot-like value and thus may be meant to satirize or ridicule CC. In all cases, independently from the affective connotation that the correspondence in the blinking may have, it can be confidently affirmed that the pleasure of imitation plays a great role in this type of responses, as also observed by Burgess (forthcoming) and Willet (forthcoming) in other (You)Tubing practices. In fact, imitation is always paired with variation, which, as evidenced also in the ‘Where Do YouTube?’ thread, is a means of communicating one’s difference while maximally attuning with the initial video.



**Fig. 55** Snapshots of blinking corresponding response.

In 71 responses the face at the camera does not blink, sometimes explicitly *not* blinking (i.e., widening the eyes and keeping them wide open, as in the two screenshots in Fig. 56). This latter case is a corresponding ‘counter-response’, in which the (You)Tuber expresses the opposite of the prompting action; its verbal homologue could be a negation, while the cohesive tie it establishes is one of opposition as in the following hypothetical verbal exchange:

- A: I blink
- B: I don’t.

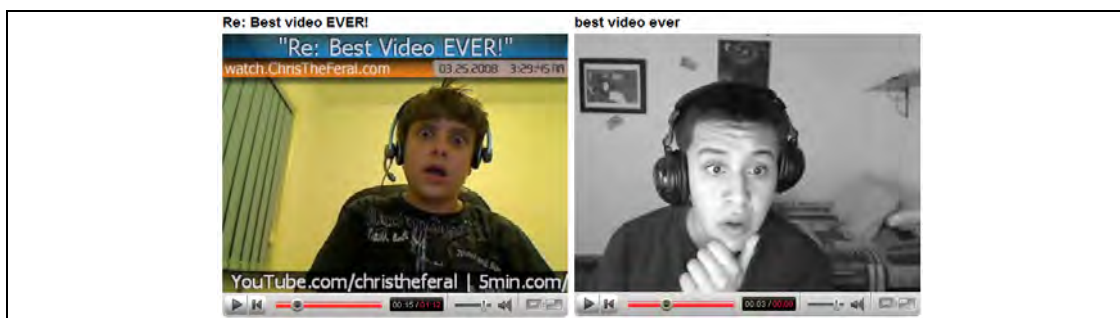


Fig. 56 Snapshots of faces explicitly not blinking.

17 responses feature the (You)Tuber variously disguised, often wearing sunglasses (as shown in Fig. 57.a). In preventing the blinking from being seen, the response with the sunglasses is evasive. Indeed, rather than negating the initial video's salient action, as in the case of the eyes wide open (i.e., the *not* blinking), the use of the sunglasses negates the *verification* of the action and thus 'evades' it. More specifically, the (You)Tuber provides an attuned response (by means of a very short silent video featuring a face-at-camera); the use of the sunglasses (i.e., a salient variation from the initial video) foregrounds the eyes, i.e., the locus of the initial video's salient action; in turn, the sunglasses prevent to ascertain the actualization of the correspondence of the salient action. Hence, the response simultaneously sets the blink as its salient prompt and responds to it by avoiding to give satisfaction to it<sup>113</sup>.

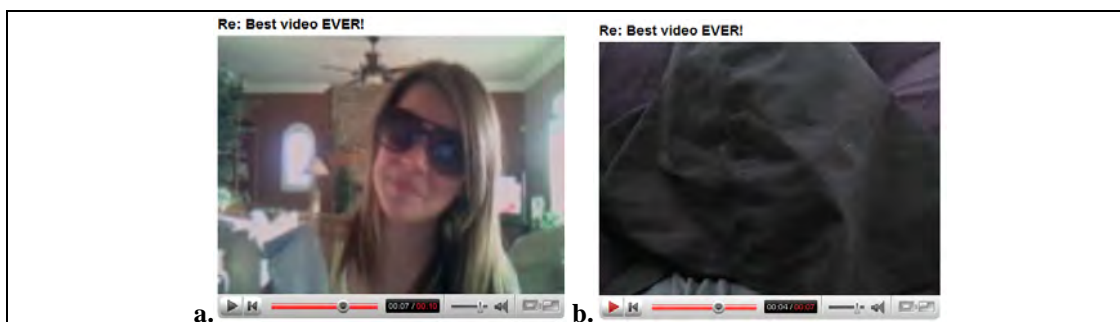


Fig. 57 The sunglasses (left) and the hood (right) as signifiers of an evasive response.

Similarly (and yet differently), another response features the (You)Tuber's face completely covered by a big hood (cf. Fig. 57.b). A voiceover says:

yes I know you can't see my face this is what makes it funny

In this case (as in the case of the sunglasses), what creates humour – i.e., 'what makes it funny' – is the evasiveness of the response, which perfectly corresponds the initial video *but* in its main action, whose verification is prevented through the hood.

<sup>113</sup> Its verbal analogue may be, e.g., the following exchange (parent vs. confrontational teenager):  
[PARENT]: what time will you be back home?  
[TEENAGER]: at some time.

What differentiates these two responses is that, while the sunglasses focus on the eyes and, by metonymy, make the blinking a salient prompt, the hood (and the spoken ‘face’) focuses on the face, i.e., on a wider entity than the eyes, so that the reference to the action of blinking is vaguer (more generic) in this response. Compared to the sunglasses, the hood zooms out the locus of the salient action. In turn, the hooded (You)Tuber’s reference to some ‘humorous’ meaning in relation to the hidden face (i.e., ‘this is what makes it funny’), communicates that the covering of the face is intentional and thus prompts the viewer to search for the source of the ‘humour’ in the response’s relation with the initial video.

In sum, in a different way than the sunglasses’, i.e., by indicating the intended effect rather than the locus of the action, also this response provides some clues on its interested prompt-response relation with the initial video (i.e., an evasive response to the blinking prompt).

While this ‘hooded’ response zooms out from the locus of the salient action to the face, a couple of responses zoom in (Fig. 58), through a close-shot of an eye blinking (i.e., corresponding CC’s blink) or staring at the camera (i.e., negating it);.

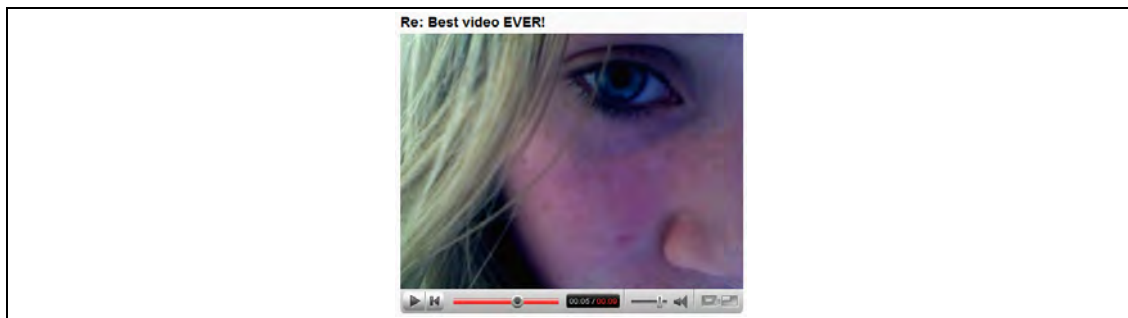


Fig. 58 Snapshot of a close-shot of an eye.

### 3.1.2 The ‘insignificant action’ as a corresponded prompt

58 corresponding responses represent the (You)Tuber doing something else rather than blinking (or explicitly ‘not blinking’). So, the featured respondents can drink (2), eat (3), chew (1), smoke a cigarette (1), simulate an epileptic attack (1), laugh (4), breath (1), make some noises, like farting (2; in this case, connoting derogatorily the initial video); they can speak short utterances (18), or make a nod (5), a ‘thumb up’ (3) or a ‘finger’ gesture (8). One (You)Tuber shows the camera a fork, then a key and a marker, while another is filmed with two lighters inserted in his nose holes.

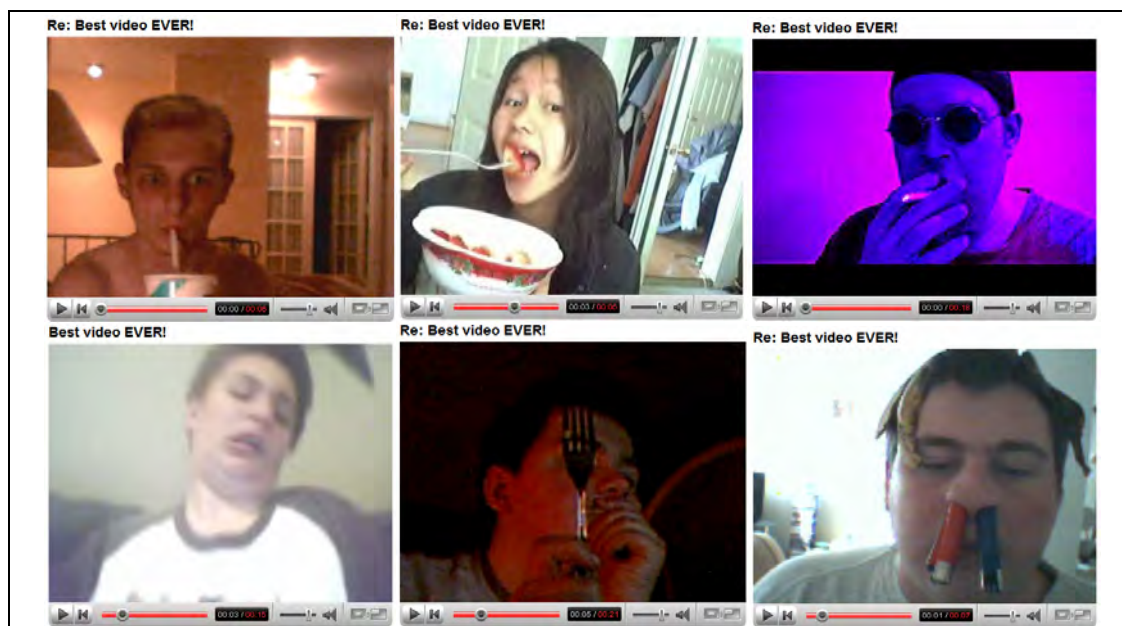


Fig. 59 Other corresponding actions.

Independently of their interpersonal meaning (i.e., whether they are intended to praise or scorn CC), all these responses correspond the blinking with another action. While the blinking ones take up and correspond the ‘literal’ ideational meaning of the salient action represented in the initial video, all these others take up a more abstract value of CC’s action of blinking, i.e., its ‘nothingness’ (referred to also in the initial video description), and respond to it with another insignificant (or non-sense) action, by – say – showing a fork to the camera. They thus establish a cohesive tie of substitution (within the semantic field of ‘insignificant action’).

In terms of interpersonal meaning, some of these represented actions are affectively neutral towards CC and the initial video. Indeed, in representing the (You)Tuber drinking or smoking, these responses do not express explicitly any clear positive or negative attitude towards the initiator. In turn, others express an affective value towards the initial video. The ‘finger’ gesture in particular – i.e., the back of a closed hand towards the camera and the third finger standing, or, in the British variant, the second and third finger jointly standing (cf. Fig. 60.a and .b) – has the value of an insult towards CC. In turn, support and approval towards CC is expressed by the nod and the ‘thumb up’ gesture – a closing greeting gesture frequently used in videos (Adami, 2009a); cf. Fig. 60.c. In these cases, besides corresponding the form of the initial video with a varied action (i.e., a gesture instead of a blink), the response functions also as a commentary (cf. the next section), in insulting or praising CC.



Fig. 60 Snapshots of finger gesture (US and UK versions) and thumb-up gesture.

Also some spoken utterances have similar (positive or negative) values, e.g., ‘I love you Chris Crocker’ (spoken by a boy wearing make up), ‘asshole stupid’ (spoken by a voiceover in a video featuring a dog), or ‘shut the fuck up man’ (spoken by a girl). Others are rather inquisitive towards the representation of the initial video, like a spoken ‘what’ (2); analogously, a (You)Tuber shows a ‘WTF?’ handwritten on paper<sup>114</sup>; these spoken utterances find their visual homologues in responses featuring perplexed facial expressions, e.g., a frown (both represented in the video in Fig. 61).

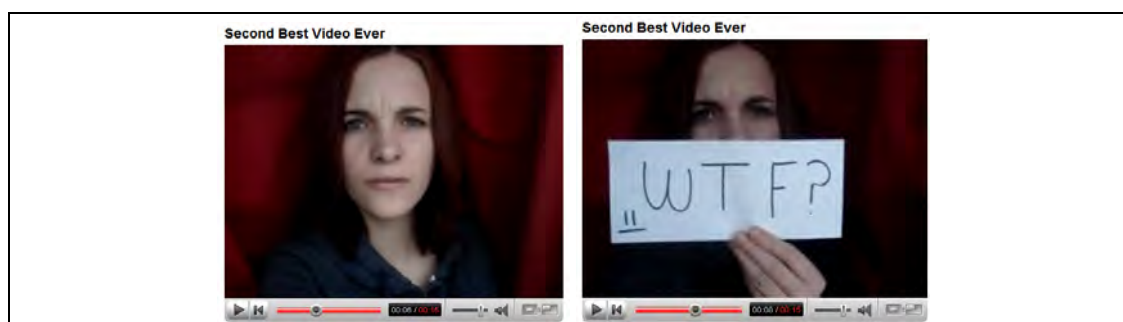


Fig. 61 Snapshot of a perplexed facial expression followed by a handwritten ‘WTF?’.

Some responses represent more than one action, such as a video in which the (You)Tuber puts a finger in his nose, then in his mouth. He then smiles and says

crap oh ouch ouch oh god I can’t stop blinking oh ouch it hurts I’m sorry [laughs] I can’t stop blinking [the (You)Tuber is blinking continuously while speaking]

This response corresponds literally the blinking of the initial video, while its formal correspondence is negatively connoted in the reiterated parroting of the action, further stressed by the spoken utterance. Moreover, in the action of ‘finger in nose then in mouth’, the response enacts a variation which conventionally communicates disgust (together with humour). This playful disgust can be interpreted as addressed to the initial video; if nothing, a disgusting representation is a provocative response.

<sup>114</sup> Interestingly, this response shares the same mode of the ‘Where Do YouTube?’ thread and could well be moved by its uploader to it, thus resulting in a perfectly attuned recontextualization.



### 3.1.3 The sexually-related corresponded prompt

Other represented actions are sexually connoted, such as a repeatedly spoken ‘penis’, or a close-shot of a female-mouth with a pierced tongue licking the lips (Fig. 62). Besides corresponding CC’s video, these responses take up also an element which is salient of CC’s (You)Tubing history, i.e., CC’s sexual explicitness in videos (both in the persona’s aesthetics and speech), which often deal with sex-related topics.

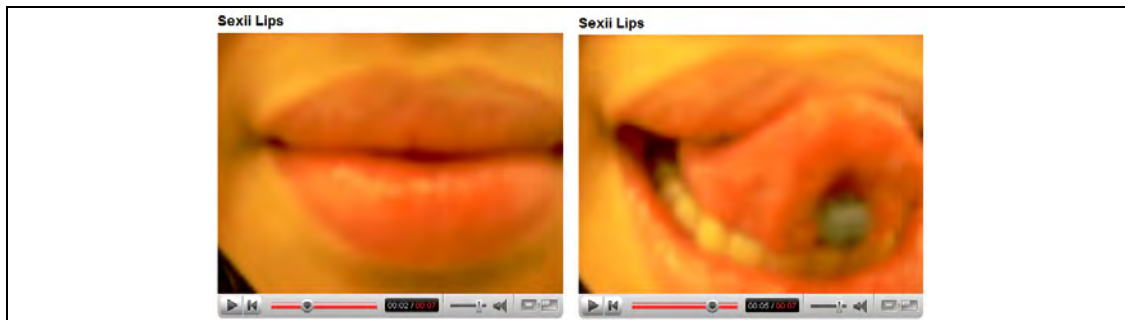


Fig. 62 Snapshots of a close-shot of lips and pierced tongue licking them.

### 3.1.4 The corresponded smile

44 responses feature a smiling face (i.e., they take up also CC’s smile as a prompt, which thus establishes another cohesive tie of repetition with the initial video), while the remaining ones portray serious mouth expressions. Smiles (which of course combine with eyebrows positions) are differentiated in the responses and each express different values, from serenity and solidarity (e.g., the one portrayed in Fig. 55 above), to defiance or even derision and scorn (e.g., in a video a (You)Tuber smiles by lifting only one side of his mouth, cf. Fig. 63.b.). Hence, even in taking up this prompt (You)Tubers modulate differently their attitude towards the initiator’s performance. A comment interprets as a scorn the smile (together with the black wig) featured in the response in Fig. 63.a:

OOOOOOHHH, I totally get it now! XD  
Way to say "fuck you" to Chris Crocker!!!  
You're fucking awesome! XD

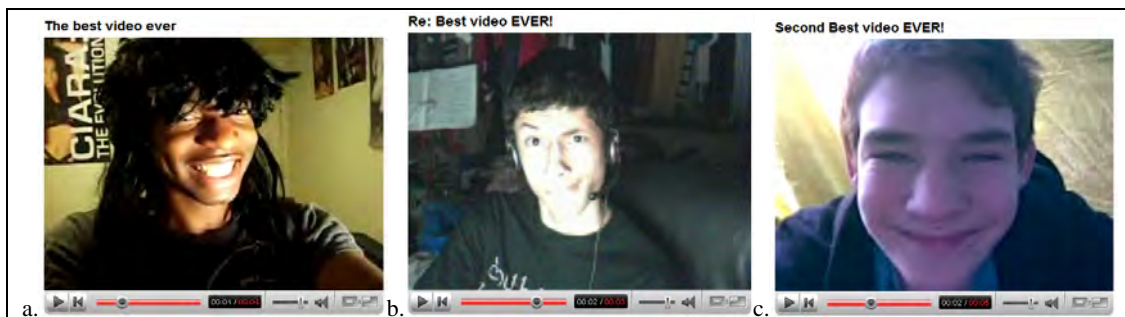


Fig. 63 Snapshots of smiling responses interpreted as a scorn.

### 3.1.5 The ‘bestness’ as the corresponded prompt

Some of these varied corresponding responses refer to the ‘Best video EVER!’ title, e.g., the one featuring a (You)Tuber saying:

I’m pretty sure that the two million thirty thousand whatever people who watched that video are gonna probably enjoy this one better. BITCH! [the word ‘bitch’ is shouted out loud]

The word ‘bitch’ takes up an idiolect word often used by CC in videos (which, especially in the queer and youth community, has not the value of a categorical insult but is rather a light-hearted mocking epithet).

The initial video title is referred to in another, more semiotically dense response (Fig. 64), which features a boy blinking, then raising his eyebrows and saying:

I think this is the best video ever

Then he blinks again, puts his tongue out, sends a kiss, smiles and laughs.

Here the respondent makes a series of actions, incoherent among them (e.g., the opposite values of the tongue out and the kiss) and thus equally insignificant. In so doing, the respondent communicates that representing more insignificant facial expressions than just blinking make a video better than CC’s.

Then he says:

hang on Chris Crocker you should go like this

and makes a ‘monstrous’ face expression (i.e., distorted mouth, eyes wide open, eyebrows lifted asymmetrically). In so doing, he suggests CC a ‘better’ (i.e., more awkward) facial expression to perform.

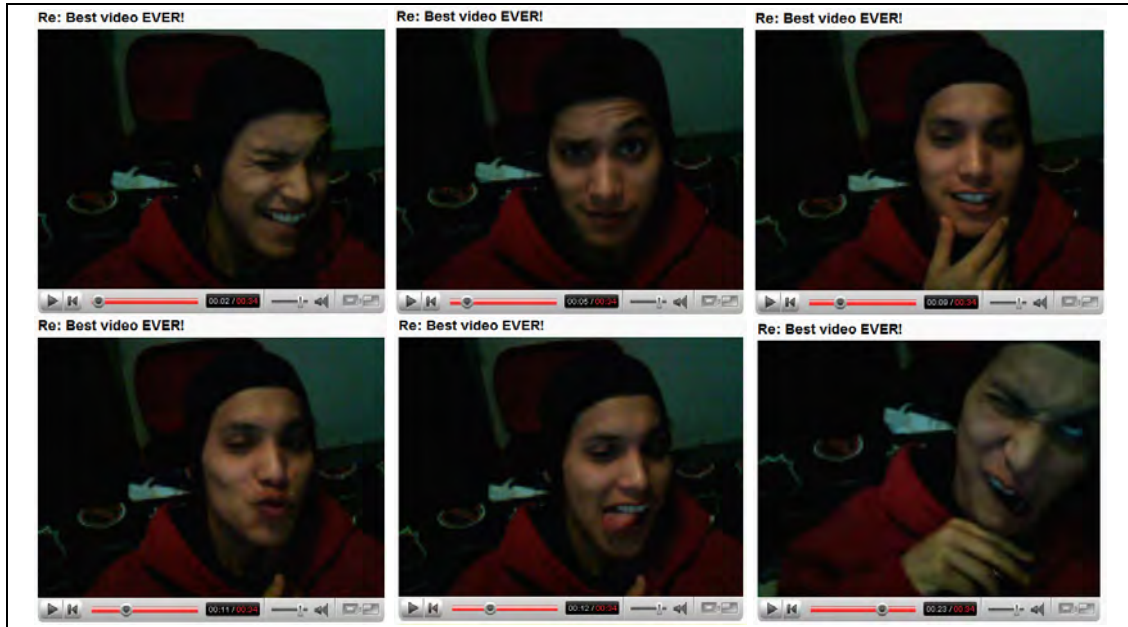


Fig. 64 Snapshots of ‘blinking’, ‘perplexed face’, ‘chin grub’, ‘kiss’, ‘tongue out’, ‘monstrous face’.

### 3.1.6 Variations in the corresponded prompts

Instead of the (You)Tuber’s face, some responses employ an avatar, variously produced, e.g., through drawings, animations or even a filmed cat or dog (Fig. 65). In one case, the dog has a yellowish mop on its head, which refers to CC’s blond hair.

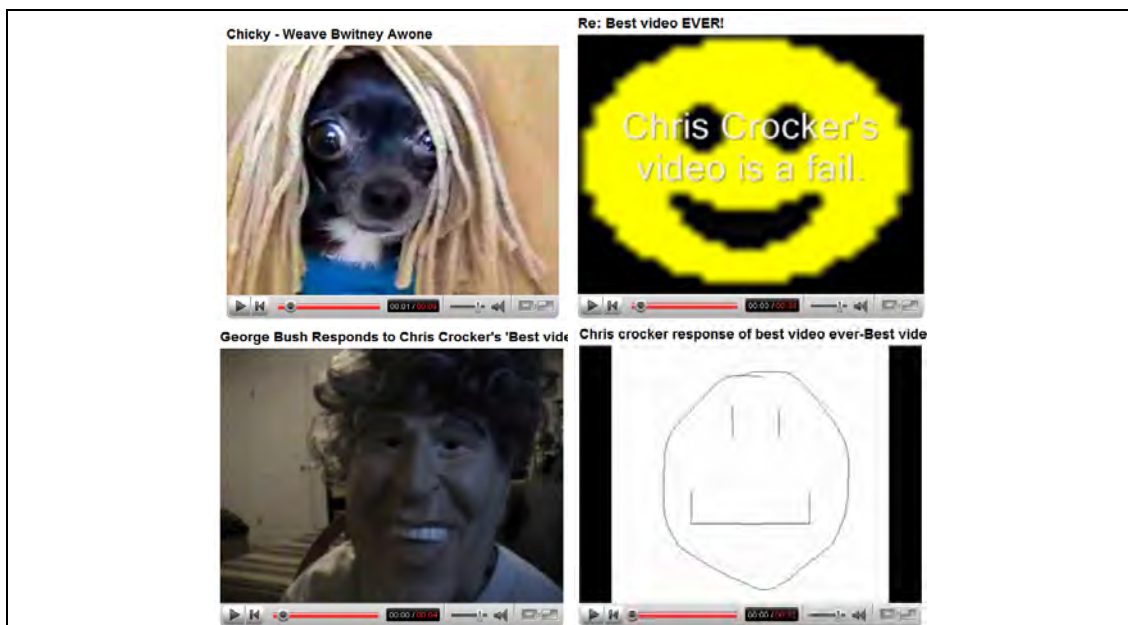


Fig. 65. Avatars: mopped dog, smiley, George Bush-masked (You)Tuber, drawn face.



In other six responses the (You)Tuber wears a blond wig (thus personifying CC) and, in four cases, also the background is the initial video's one, i.e., a whitish blanket (Fig. 66). Wigs and blanket establish further cohesive ties of (varied) repetition with the initial video. Besides, these responses – and the ‘mopped’ dog one – parody CC and the initial video and thus fall also into the ‘original spoof’ category (Section 3.4).

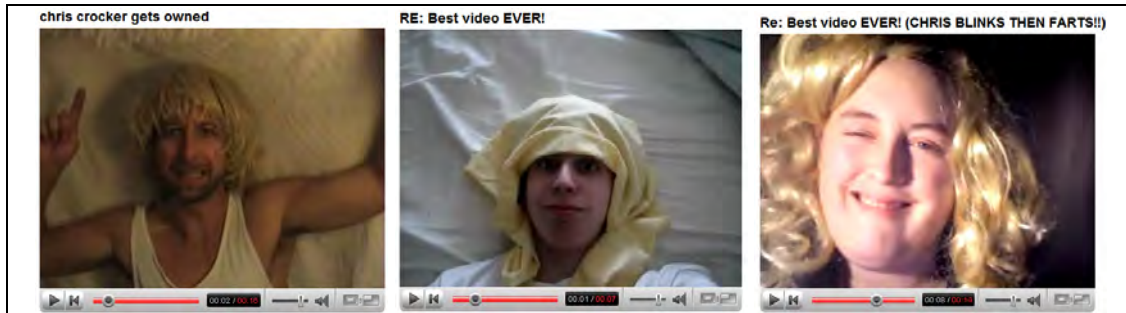


Fig. 66 Snapshots of blond wigs and blankets in the background.

Other variations in the responses include the type of shot (so that, a response reverses the shot upside down) or colour effects (either in black and white or with high contrasts or cartoon-like effects).

Four responses feature more than one face looking at the camera. Here (You)Tubers may be shot separately or jointly, each doing or saying something in their turn. The latter case is exemplified by a response (Fig. 67) featuring three boys, keeping silent for some seconds. At some point, one of them addresses the initiator by saying ‘hi Chris Crocker’, the second uses CC’s characteristic phrase ‘bitch, please’, while the third rephrases CC’s famous video, by saying ‘leave Chris Crocker alone’. The three utterances are pronounced into a *crescendo* of laughter.

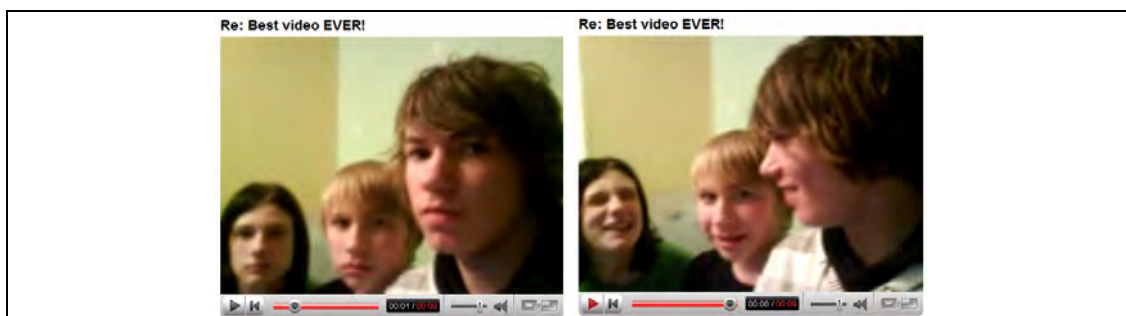


Fig. 67 Snapshots of the three faces’ response.

This response, as many others in the corpus, is ambiguous in its affective value towards the initiator. Indeed, if taken literally, it (1) addresses CC, (2) quotes one of CC’s famous interjections, and (3) pleads for CC’s freedom from bashers by misquoting one of CC’s famous videos. The laughter gives a clue that the utterances should be taken ironically, however, it may also be an unintentional outburst (as

sometimes happens when friends enact a script together). Without the video description, which in this case expresses insulting criticisms of CC ('Response To Chris Crocker. God, he's such a tit. SUCK SALAD BITCH!'), it is difficult to determine with certainty the authors' attitude towards the initiator.

In many cases, indeed, a clear interpersonal meaning cannot be assigned straightforwardly in these responses. First of all, intertextual reference and imitation are practices which are enjoyed in themselves by (You)Tubers; hence there needs not necessarily be a clear 'pro/against-CC' value in their intentions, or, better, they may not be interested in manifesting an explicit attitude towards the initiator. Furthermore, explicitness is often not sought for in videos and, even more, ambiguity or vagueness are intentionally communicated. Indeed, it is often the case that videos do not bother to make a message clear and explicit. In this type of interactions the purpose of communication lies more in the performance, in the very act of communicating (i.e., the phatic function), in tuning in with the initial video while producing variation – and thus communicating one's difference – rather than in expressing a clear interpersonal meaning. In most cases, the overall 'fun' is in the performance itself, in the pleasure of playing with the medium and semiotic modes in creative ways. Analogously, it is not the understanding of the interlocutor's intentions which matters in video-interaction, but rather the possibilities, the prompts that the interlocutor's text opens to transformation, to make new texts. As *YouTube* habitual viewers have surely experienced, puzzlement is a very frequent effect produced by *YouTube* videos. And part of the fun resides in attempting at giving a personal interpretation to them; i.e., in solving the puzzle which has no unique solution (and may well have no solution at all).

<b>Blink</b>	<b>Smile</b>	<b>Action</b>	<b>Mask/avatar</b>	<b>Wig</b>	<b>Blanket</b>	<b>Others</b>
54	44	58	17	6	4	11

**Fig. 68 Prompting Elements taken up by the corresponding responses in the thread.**

All considered, these corresponding responses take up the ideational meaning of the initial vide with no reference to its interpersonal one (i.e., to prove CC's celebrity). The blinking ones take up the literal meaning, while the others take up another – more abstract or hyperonymic – level of meaning, i.e., the insignificance of the blink.

The wide range of variation deployed in the corresponding responses (Fig. 68) testifies even more to the fact that the practice of video-interaction works on an understood challenge for differentiation-within-attuning (also observed for the 'Where Do YouTube?' thread), i.e., a strive to deploy the maximum extent of variation stemming from a given kernel, which is here even more limited than in the other thread. Indeed, here the very short video and the very few elements represented in it constrain even more the range of possibilities for variation. Also here, when deploying the same multimodal pattern in the response, respondents strive to differentiate their contributions, while maximally attuning with the initiator.

Significantly, this attuning does not coincide necessarily with an expression of support or approval towards the initiator (in the same way of the handwritten insult ‘WELL RNT U A NOSEY BASTARD’ in the ‘Where Do YouTube?’ thread; cf. Chapter 5, Section 3.1.2). The interpersonal meaning is often ambiguous and, as discussed here, it often seems not to be even relevant. What really is at stake in these corresponding responses is the pleasure of transforming and taking up the initiator’s signs as resources for a new representation.

In sum, this variation-within-attuning is a newly-shaped form of cooperation (differently conceived than Grice’s one) which hinges on a differentiated participation rather than on understanding (or, even more, supporting) each other. ‘Support’ is given to the initiator by merely posting a response to the initial video (and thus contributing to enlarge the thread and make the initial video more visible), even when insulting the initiator (by, e.g., featuring a finger gesture).

### 3.2 Commentary responses

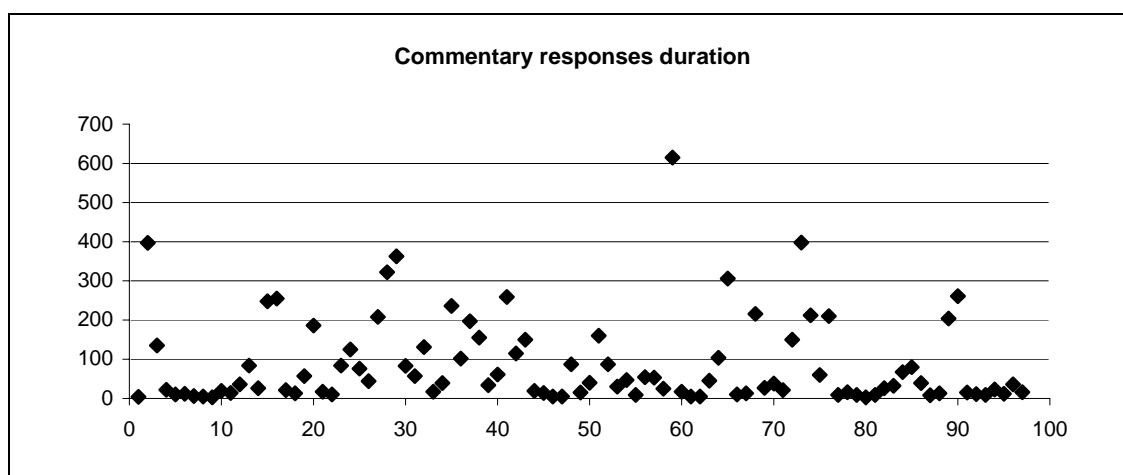


Fig. 69 Distribution of commentary responses according to their duration: 97 videos; mean 87”.

Other responses (97) comment or evaluate the initial video or its author. Most of them deploy the typical features of videoblogging (i.e., the (You)Tuber more or less randomly ‘ranting’), either praising or criticizing – at various degrees, from arguing, up to employing ‘hate’ language against – the initial video or its author. As shown in Fig. 69, these responses are generally longer than the ‘corresponding’ ones (Fig. 54). Like the corresponding ones, also these responses take up specific elements of the video or of its featured character and make them the salient focus of their comment.

### 3.2.1 The initial video commented

#### *a. The video 'bestness' as the commented prompt*

Those who praise the initial video can refer to and acknowledge its title; in mentioning it, they establish a cohesive tie of repetition; so, e.g., a response features a boy sat in a corner, looking at the camera and saying

hey, Chris that was THE best video ever

Some praise the 'bestness' of the initial video against its detractors; in a response a girl faces the camera (while another one sits behind her on a sofa) and says:

ok shit it's recording ok anyway I just wanna say something to all the fucking haters out there ok you need to stop ok if you weren't to take the time to watch the video and like waste your time talking to shit that is really pathetic you need to get a life. You know that was the best video ever ok Chris Crocker is the best video ever ok shut the fuck up all the stupid haters kiss my ass cause I know you want to and I'm sick of it I'm sick of all the haters fucking shoot about Chris Crocker ok cause I bet you anything that he is better than you you're just low piece of shit so stop wasting your time if you don't like Chris Crocker then stop watching his videos and leave him alone ok

Others comment negatively the initial video; so, a girl starts laughing and says:

that was really stupid video, pointless, just like this video, pointless, pointless  
pointless

In pairing her response and CC's video as both 'pointless' (i.e., 'just like this video'), the respondent expresses her selection of 'pointlessness' as the criterial aspect for her interested actualization of the prompt-response relation with the initial video, analogously to the 'nothingness' corresponding responses discussed earlier (3.1.2).

Some enact a performance which implies evaluation, as in the video in Fig. 70, in which a boy (filmed in profile, his face illuminated by the screen light) says:

hey, what's on You Tube today mmm mmm ah Chris Crocker

The (You)Tuber watches the screen with a perplexed facial expression; the camera shoots the screen playing the 'Best Video EVER!' and then the (You)Tuber with an astonished expression saying:

what the hell... aaah!

The image disappears and, while the soundtrack plays 'Titanic' (C. Dion), the following typing appears on the screen:

this poor young boy has recently died after seeing something so crap. R.I.P.

This respondent enacts his reaction to the watching of the initial video. In so doing, the video takes up a well known ‘meme’ on *YouTube*, i.e., the ‘reaction to 2 girls 1 cup’. ‘2 girls 1 cup’ is a porn fetishist film, and hundreds of videos are posted on *YouTube* showing (You)Tubers (or other unexpected agents, e.g., their dog) watching the screen – which emanates noises characteristic of the porn film (thus implying that the screen is featuring the ‘2 girls 1 cup’ film) – and enacting a varied reaction of disgust (in this way they also scorn the prohibition of showing pornographic content on *YouTube*).

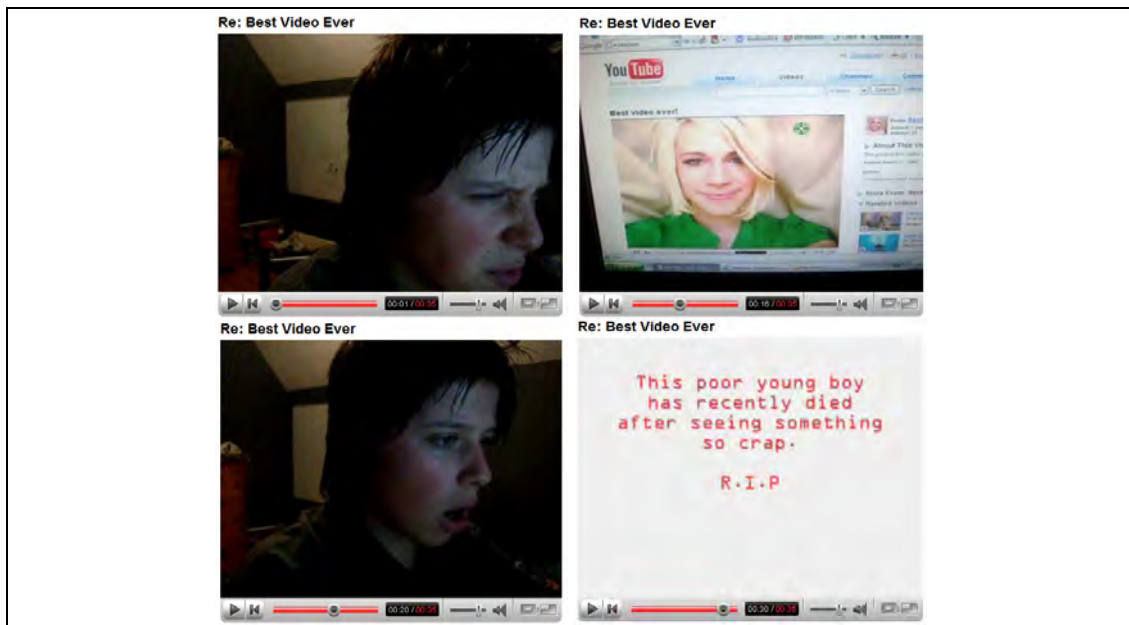


Fig. 70 Screenshot of a response embedding the initial video.

Playing with genres and intertextual references is a widely practiced on *YouTube*, which produces enjoyment in referring implicitly to elitist knowledge (i.e., only experienced (You)Tubers can get the intertextual reference), as discussed in Sections 3.3.2.b and 3.4.2.b.



Fig. 71 Snapshot of the Guy Fawkes-masked (You)Tuber who comments on CC not being *Magibond*.

So, among the video detractors, some refer to *Magibond*, a quite popular Japanese girl who often videoblogs staring silently the camera; a Guy Fawkes-masked (You)Tuber (Fig. 71) says:

I am not going to judge but you are not no 21 year old Japanese girl named Magibond ... you'll never be Magibond

In so doing, these respondents point to CC's copied performance and lack of inventive and thus undermine the 'Best video EVER!' title of the initial video (for *YouTube* fostered equivalence between the 'best' and the 'first' in representing something, cf. Chapter 4, Section 2.4.7). In sum, they take up the silent staring of the representation as a prompt to which to respond by pointing out its non-novelty, and, in the same time, they express intertextually their experienced status as (You)Tubers (i.e., we know all important genres here and we can point out similarities and differences); cf. other examples of intertextual reference in 3.3.2.b and 3.4.2.b).

*b. A background element as the commented prompt*

Commentaries can take up a background element of the initial video and make it the salient prompt to which they respond; for example, a girl focuses on CC's make up:

Chris Crocker I just wanted to say that your make up is absolutely beautiful the way that you use the shades and the colours I think that it is fabulous. Just keep the make up the way you are cause in the best video ever that is possibly the best make up that I have ever seen done.

Another commentary makes salient the absence of an element, rather than a background element. It features a boy's eye at a close-shot and a voice saying

hmmm I'm disappointed about the lack of ninjas in this video ahhh

Judging from its title ('Re: 悪作剧2吻 They Kiss Again Episode 14 Part 4 of 9'), the response was moved from another prompting video and recontextualized as a response to CC's. A comment questions the video relatedness with its title:

what does this video have to do with the title??

The respondent replies to the comment as follows:

what are you talking about!?!@ surely you meant what does it NOT have to do with the VIDEO. There is not ONE ninja in the whole video. Did anyone watch this and not notice that! I don't respond to titles, just videos, FYI.

This reply is particularly interesting to our case, in that it substantiates the heuristics of an interest-driven prompt-response relation. Any semiotic act in video-interaction is taken as the chance for a creative response, often in terms of selection and

recontextualization of pre-existing semiotic material, according to the interactant's interests. The final remark 'I don't respond to titles, just videos, FYI' (= for your information), makes explicit that the commentary in the video is to be intended as addressed to the responded video and not to the one indicated in the 'Re:' title. In other terms, after watching CC's video, the (You)Tuber has thought that his previously made response (to the Japanese episode, as expressed in the automatically generated title) could fit more aptly – according to his interests (which may include gaining more visibility) – as a recontextualized response to the 'Best Video EVER!'. This recontextualization practice is analogous to the rearrangement of files and folders on a computer according to a new interested relatedness rationale.

### 3.2.2 CC's character as the commented prompt

Many commentaries evaluate and comment CC, rather than the video<sup>115</sup>. Also in this case, they can take up various (foregrounded or backgrounded) elements as their salient commented prompt, as illustrated here-below.

#### *a. CC's past production as the commented prompt*

Videobloggers often mention some of CC's past video production either praising or scorning it. A very frequent commented video is CC's most famous 'Leave Britney alone' video. A respondent asks for votes for 'Chris Crocker's Leave Britney Spears Alone video in the YouTube awards', because 'he deserves it'. Another says:

hi Chris this is Mina Hodgkin from Switzerland I bell on your videos one day by accident and I love them I think you're a very very crazy person and the video that makes me laugh the most was the 'girls like their gays' it's really right... the Britney one too you were crying for her I think you're right cause she had so terrible things... I like you very much... hope my English is not too bad.. bye bye

Like the praises to the initial video, also these can support CC against haters; cf.:

ok so this video is for Chris Crocker and if you don't like him I don't care and if you don't like me I don't care so walk away walk away walk away so whatever

After saying that, the girl starts the music player, but has a second thought, and adds:

and the song is Britney Spears' so that's why I'm doing this because he likes Britney Spears I don't know if he likes this song he's probably the most loyal Britney Spears' fan I'll sing this song for him now

She then sings the song which is played in the background. This respondent not only

---

<sup>115</sup> Of course evaluations to the initial video and to CC intertwine in responses; cf. 'Chris you're like I think you're awesome and the serious videos you've done are really great but this one is like I don't know subtle I don't know if this is the best one ever but all right if you say so' (spoken excerpt).

praises CC but pays also homage to CC by singing a song of CC's favourite singer.



Fig. 72. The 9/11 commentary to CC.

Those who criticize CC can take up some of CC's most provocative semiotic acts. For example, a video response (Fig. 72) starts by embedding (i.e., a direct-quote) an excerpt of a CC's videoblog, which outraged many, where CC is saying:

I can't think of 9/11 when Britney is going through what she's going through.  
Britney is a national treasure, who cares about 9/11

The response concludes with a slideshow of photos of the 9/11 attack to the World Trade Center, while the names of the casualties scroll on the screen.

*b. CC's idol (Britney Spears) as the commented prompt*

Some address CC and comment on Britney Spears, CC's idol (and the topic of CC's most famous video). For example, in a response, a boy says:

hey, as I stated in my previous video, Chris, here's a hobby for ya

He shows a digital game DVD to the camera and he adds:

here is a hobby and stop worrying about Britney this game just came out last week  
get yourself play this game and stop worrying about Britney she can take care of  
herself, her and her sister need to know the norm of birth control it's as simple as  
that

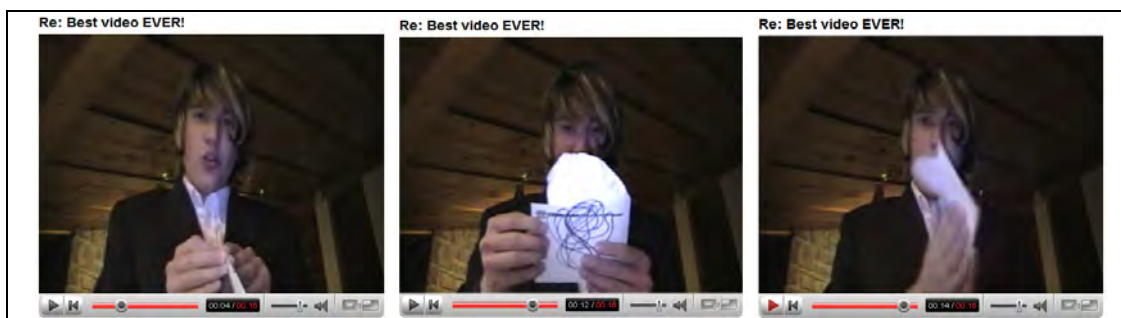
*c. CC's fame as the commented prompt*

Some respondents point to CC's ambitions for fame. A (You)Tuber (Fig. 73) paraphrases a typical 'award nomination' situation; he opens an envelope and says:

and the award for the most likely to overdose on his own ego goes to ... Chris  
Crocker for his egotistical ramblings

Then he claps.





**Fig. 73** Snapshots of the nomination-enactment response.

Another (You)Tuber addresses *Chris Crocker* directly and says:

to be honest you're such a attention whore, you're just a fucking idiot. Peace

Analogously, another one says:

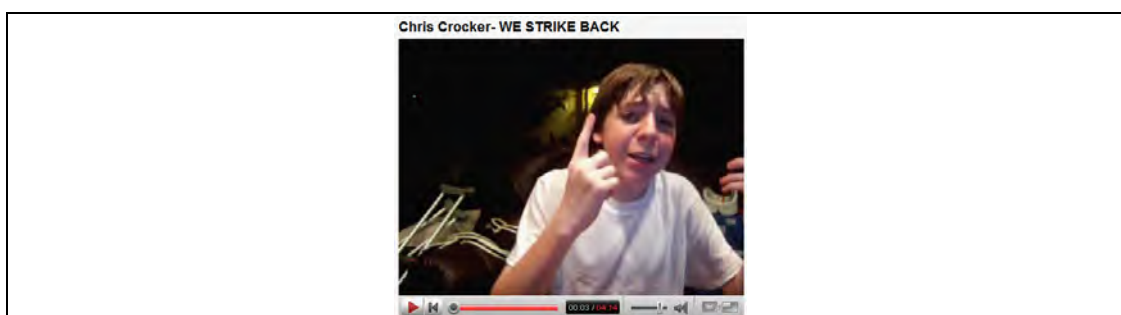
hey Chris Crocker it's your fifteen minutes of fame calling and I wanna refund

The 'time refund' claim refers to a concept which is frequently expressed by comments criticizing videos, e.g., 'you wasted [no.] seconds of my life; I want them back/refund' (with the '[no.]' filled with the duration time of the video in question). Indeed, another response, featuring a 'finger' gesture, has the following description:

5 seconds back! you fag!

CC's celebrity element can be taken up and addressed to the viewers. So a boy says.

ok let me ask just one question: did you click on this video just because it had Chris Crocker name on the top?



**Fig. 74** Snapshot of the index finger indicating the video title.

While saying 'Chris Crocker name on the top' he raises his finger onto his head (Fig. 74), so as to indicate the title of the video, which is 'Chris Crocker- WE STRIKE BACK'. Then he adds:

You know seriously if you're that obsessed don't. Quit being obsessed about him if you hate him you know you don't need to be that obsessed about him. If you like him whatever do your things stay out of my business.

He then announces his intention to show 'how Chris Crocker is a fake'; he shows the first video in which *Chris Crocker* talked about Britney Spears (by turning the camera to the screen) and discusses how CC has set up the whole 'Britney Spears' fanaticism just to get attention. In so doing, and in using CC's name in the video title, the (You)Tuber exploits CC's celebrity.

d. *CC's sexual orientation as the commented prompt*

Among those who criticize CC, many take up and comment on the character's sexual orientation. Unsurprisingly, yet quite appallingly, homophobic responses are numerous. For example, a response introduced by a typed 'MY REPLY...' shows CC's face taken from one of CC's previous videos, while the soundtrack sings 'everyone has aids', followed by a typed gay-related insult 'YOU ARE A GIGANTIC FAGGOT' and a slideshow of gay-hate related images (Fig. 75).

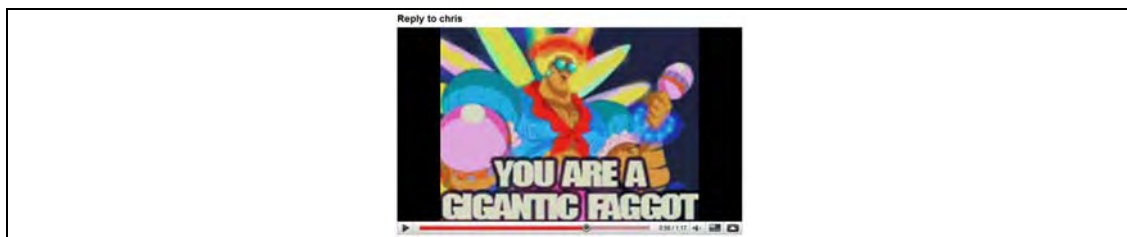


Fig. 75. A gay-hate commentary.

Another boy faces the camera and says:

Chris you are a gay gay gay gay idiot I hate you you're wasting our time

In turn, some criticize *Chris Crocker* but distance themselves from gay-haters, like a young boy who refers to a previous video of CC against gay bashing (thus taking up an element of CC's (You)Tubing history, rather than the initial video) and says:

I saw Chris Crocker's video the other day about gay bashing on YouTube I don't now how long it's been out. I don't like gay bashing. I talk about Chris Crocker. Chris Crocker I talk about you so bad and I hate you and I want you to die but that's different see I don't like Chris because he dresses like Britney Spears and he's transsexual damaged weirdo I have gay people in my family so I'm cool with gay people but I don't understand why he's got to get on YouTube and make such a big scene about it and all these other gays now posting videos go like 'yeah Chris thank you fabulous that you did it' we do not like gay bashing as much as you love wanna think about it we don't are gay bashing at least I am not I'm talking about Chris Crocker not about gay people Chris I hate you. die. I'm not talking about gay people about like 'Chris you're a homo' that would be making fun about gay people or 'Chris you and your boyfriend are stupid' Jesus it's the way you go. I'm

not saying that I'm saying that you're just stupid for crying on YouTube and then 5minutes stands and you wanna rug me do you think I'm damaged. You're making videos like 'eat my corn hole' and 'the bitch bell' oh my god I know people are gonna call and getting mad at me about my video I don't care because I don't like Chris Crocker till the day I die I do not like Chris Crocker none in my family likes Chris Crocker as much as you wanna think it's not gay bashing anyway I'll probably talk to you later so if you guys wanna leave comments.

Significantly, although distancing from 'gay bashing', his criticism towards CC is in fact intermingled with transgender-hate (cf. 'he's transsexual damaged weirdo'), beside criticizing gay people who praise CC as a spokesperson for gay-related issues.

Homosexuality is a 'hot' and thorny issue for respondents, so, among those who support CC, males often feel the need to express their heterosexual orientation; e.g., a boy videoblogs praising *Chris Crocker*, while saying:

he's funny... i'm not gay but gays are cool

### 3.2.3 The comment as a chance for further representations

Some responses take the chance of commenting the initial video so as to develop a topic which interest them. For example, while praising the initial video, a girl addresses CC a specific invitation, presented as a challenge:

hi again Chris two words for you simply amazing amazing all you did was blink twice and flash that wonderful smile of yours and you got all those people to view your video I was just wondering Chris you know do you think that you'd have that much draw if it was just a voice instead of a video what do you think Chris I wanna challenge you no mean a bad challenge cause I challenge positive because I want you to succeed I sent you an invitation to be a guest on Crossfire Radio you didn't answer.

Then she introduces her radio and promotes it ('it is uncensored'); she challenges CC ('how many listeners could you get on a radio station with one appearance?'), while warning CC that 'of course we don't pay you' since 'we survive on donation'. The video is quite long (6'38") and, by challenging and inviting CC to feature in her radio programme, the videoblogger takes the chance to introduce and promote the radio itself; it is again a variation on the actualization of the prompt-response relation which is driven by the respondent's interest. Indeed, she takes up the initial video (its featured blinking and number of views) as a prompt and uses it in a flattering and challenging way according to her interests, i.e., to promote her radio programme<sup>116</sup>.

---

<sup>116</sup> This pattern is by no means rare in other forms of interactions. The same praise-challenge-self-promotion occurs often also in formal situations, as, e.g., in academic debates, when scholars intervene at conferences starting by praising the presenter's work, then asking with a humble tone about its validity in a different situation, which, not incidentally, coincides with the questioner's research, which is thus introduced and presented (i.e., promoted).

A (You)Tuber praises CC's ability to use the 'system', and, by so doing, he takes the chance to discuss and criticize the 'star-system' on *YouTube*:

this is a video review and response to Chris Crocker best video ever [he makes the inverted commas gesture]. I've been watching Chris Crocker videos quite a bit and they give me quite a laugh sometimes but this one I thought it did actually I saw some of the comments some people like oh time waster but I think that as well but in a way I think it was actually a good point to prove that he's such a massive celebrity on YouTube that he does a one video that's about ten seconds long just of him going [and he stares and blinks] and he gets one million six hundred five thousand eight hundred and 79 views. It's really mad actually and it shows you what if he has well he shows you the power which he has on YouTube which is absolutely mad. And yeah so that's just a review to his video best video ever and I'll probably be doing some more Chris Crocker reviews later today.

Analogously, a respondent discusses *Chris Crocker* and takes it as the chance to introduce the more general topic on the debasement of contents on *YouTube*:

I dont have TV I hate TV so I found YouTube... but the time passed... videos today... fucking Chris Crocker

Then he reads the description of CC's initial video and adds:

fuck. I don't know

He reads the tags, shows a snapshot of the initial video and says:

what the fuck is this video doing in the most viewed it's not even a video it's a fucking picture. I don't know, this is worst than my videos and I'm not anyone [...] people are trying so hard to make videos... it's getting worst and worst and you know it's not YouTube it's fucking you that click on that fuck

In sum, these responses take CC as an interest-driven prompt to discuss a wider topic which concerns them, i.e., the celebrity system on *YouTube* which, according to them, disregards valuable content and fosters its debasement.

---

In sum, responses can comment either on the initial video or on CC. They can take up an element and make it salient, selected from the video (i.e., the make up) or from the character (CC's fame or sexual orientation), and/or they can refer to CC's (You)Tubing history (by referring to some of CC's previous videos and their topic, e.g., Britney Spears). They can comment positively or negatively and can widen the criticism and blame viewers for giving CC an undeserved celebrity status. Out of 97 commentary responses, 24 evaluate positively the initial video and/or CC, 60 express a negative evaluation, while 13 are either neutral or ambiguous towards the quality of their evaluation. Criticisms span from polite arguing against CC up to offending CC and employing hate language. Although having the power of denying responses, CC

has kept all of them (even the most insulting ones) in the thread. This testifies to the fact that CC is interested in gaining visibility by means of the (You)Tubers' participation, no matter if participation involves insulting and even threatening CC (cf., one of the a.m. excerpts, in which a (You)Tuber wished CC to die).

In many cases (such as when they take up CC's past deeds), these exchanges would not be normally considered as coherent, since they do not follow Grice's relevance maxim. When flouting this maxim, it is generally retained that one uses hedging devices such as 'by the way' or 'totally unrelated but', i.e., phrases and markers which signal that the speaker is going to break the maxim of relevance. In absence of this, the interlocutor can question the relevance of the response (i.e., 'what does this has to do with what I've just said?'). In this thread, respondents never hedge their flouting of Grice's cooperative principles; however, comments never question the 'incoherence' of the exchange<sup>117</sup>. It is understood that viewers are required to search for the relevance in the exchange by activating a series of implicatures. Significantly, this search does not follow Sperber and Wilson's relevance principle (cf. Chapter 2, Section 1.3), in that the relatedness is not retrieved according to the principle of achieving the maximum effect (information) with the minimum effort (inferential work). Implicitness is greatly practiced in responding (such as in the commentary citing *Magibond*, discussed above), which pairs the strive of puzzling the viewer. The enjoyment of intertextual references and implicitness assimilates video-interaction to many elitist forms of communication, including some artistic ones and those used among elitist communities (like academics, for example). Rather than in the content itself, the enjoyment in these exchanges resides in getting the – at times trivial – implicit reference through a maximum inferential effort (i.e., which can be accessed only by insiders). This affective characteristics of video-interaction is found – and even maximized – in parodying responses, analysed in the next two sections.

---

<sup>117</sup> Note that the commenter to the response lamenting the absence of ninjas in the initial video does not question the relatedness of the exchange but that of the video content with its title.

### 3.3 Remakes: From remix to parody

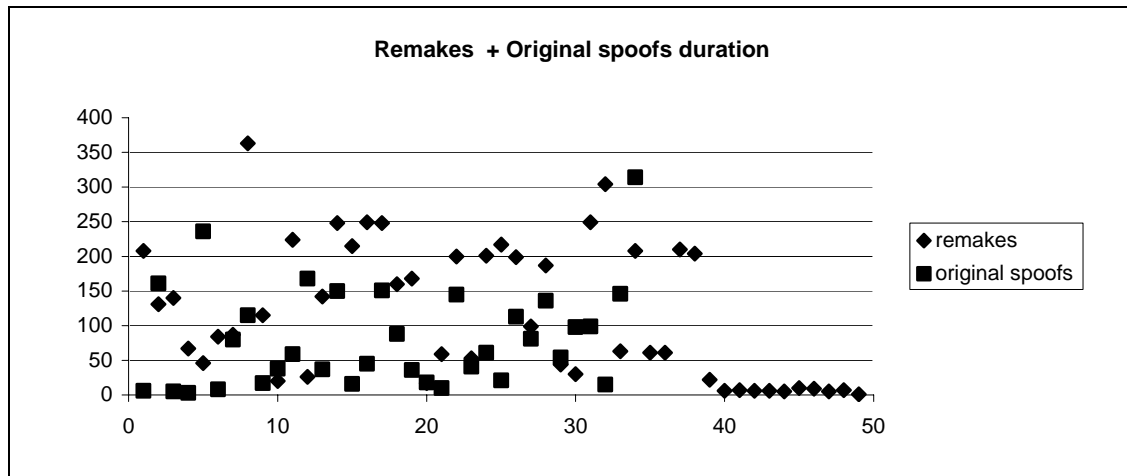


Fig. 76 Distribution of remakes and original spoofs according to their duration. Remakes: 49 videos; mean 116"; original spoofs: 34 videos; 81.5".

Fig. 76 shows the duration-distribution of two types of responses: the 'remixes', i.e., videos which reuse or embed CC's original material, and the 'original spoofs', i.e., videos which use new material to 'imitate' CC (or the initial video). The remixes are analysed in the present section, while the next one is devoted to original spoofs.

49 responses post remakes of the initial video by using its very same semiotic resources. They can simply repost CC's video, they can post an edited version or they can embed it in their original material. Some remix in the form of a mashup other CC's videos, thus, analogously to the commentaries, the remakes can take up either (or both) salient elements of the initial video and/or of its character. Let us consider each sub-type separately.

#### 3.3.1 Recontextualization of the initial video

##### a. *Reposting*

Three responses repost the initial video as it is. One of these posts a zero-second video, with a snapshot of CC's video; the uploader's channel (*ViewsTester*) posts other people's videos with the declared aim of testing *YouTube* indexing system. The response title is 'Best video EVER! The Blink!', while its description reads as:

The best part of this video is the blink of the Crocker  
Cris has 23-24 videos with over 1,000,000 views, you can check here:  
[http://uk.youtube.com/profile\\_videos?...](http://uk.youtube.com/profile_videos?...)  
his "leave britney alone" video has had over 18,000,000 views.  
Also here is a link to an article about him:  
<http://www.msnbc.msn.com/id/23748983/>  
-----

I was asked why I had "fag" in the description here, so here is the awnser I gave:  
 Well I don't hate him really, I know a lot of people do, so I wanted to see if the word "fag" would ge indexed on youtube's search...  
 Here is actually my opinion about gays, I don't care what people do in their bedroom, it is their own personal stuff. I do however not really want to hear about what people do in thier bedroom. I have a nabour who insists to tell me how to have sex (he is not gay BTW), and I really don't want to hear it. What me and my wife (maried for over 7 years) do in bed is none of his business.

As can be inferred by the description (and its further specification on its use of the keyword 'fag'), the posting of this remix has no intention either to praise or detract CC. Simply, the respondent uses CC's initial video to his purposes. In our terms, the initial video's large number of views (and its design intended to achieve them) functions as a prompt which is taken up by a respondent who is interested in monitoring the popularity-indexing system on *YouTube*.

Other two responses post CC's initial video with a modified title, 'Worst video ever', thus counter-responding to the initial video in reversing the meaning of its title. One of these responses is made by means of a camera filming the initial video playing on the pc screen, so that, when viewers play the response, the 'flv' bottom bar on the Website is exactly below the bottom bar of the filmed video (Fig. 77.a.). The other response is a mere repost which, however, adds a further gay-hate meaning in its description: 'GAY ASS HELL!!!' (Fig. 77.b.).

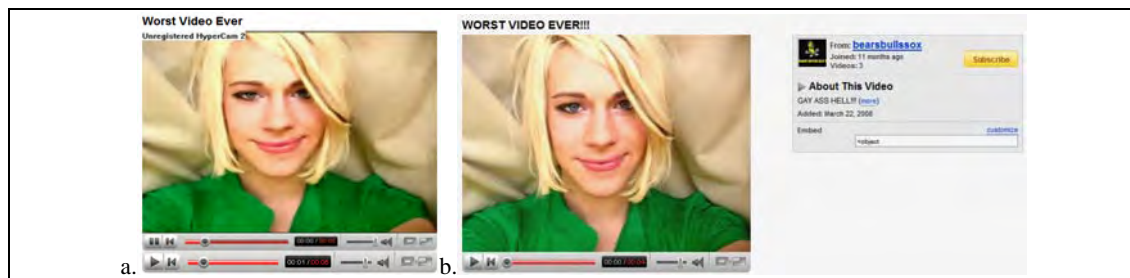


Fig. 77 Snapshots of the two 'Worst Video Ever' responses.

#### b. Editing

Nine responses post a modified version of the initial video. One starts with an introductory opening typed on screen: 'this is basically how boring chris crocker is now....', and then it plays the initial video in loop, by changing colour effects and adding a 'heavy metal' soundtrack (Fig. 78).



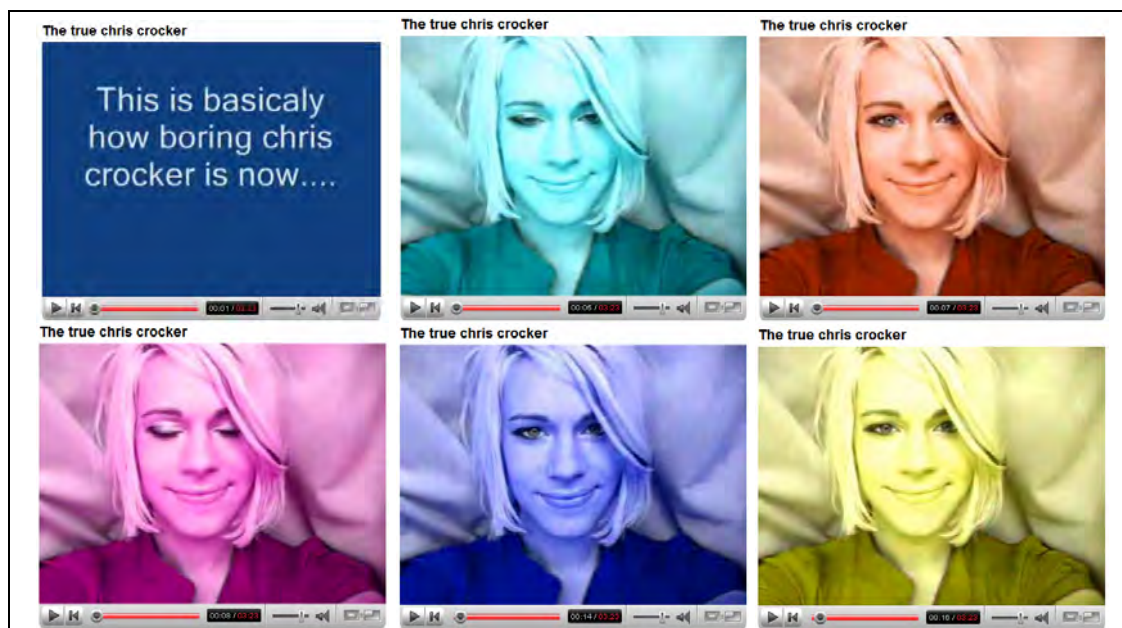


Fig. 78. Snapshots of a remix response.

Fig. 79 shows other edited versions: *SuperMario*'s<sup>118</sup> angry face is superimposed onto CC's (a); paint-brushed spots are added to CC's face (b); CC's mouth is enlarged in a monstrous smile and CC's eyes are painted in red, with lions groans in the background (c). These remixes more clearly parody negatively the initial video.

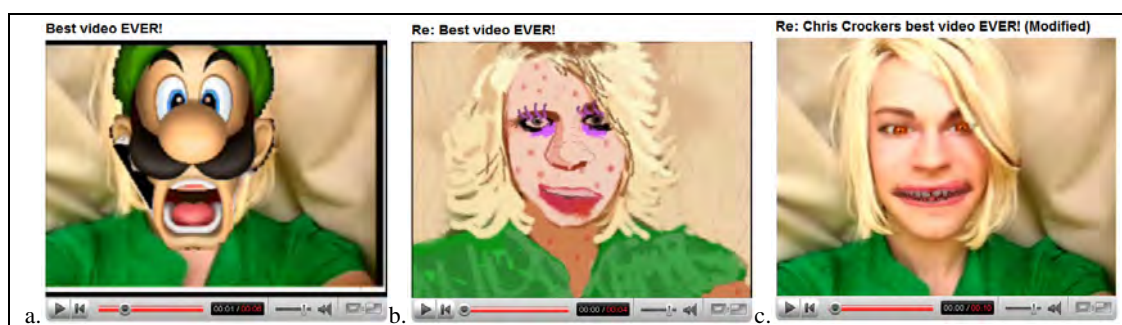


Fig. 79 Edited remixes: SuperMario, spotted face, red-eyed and enlarged mouth CC.

### c. Embedding

Responses can embed the initial video in original material, this way using it as a direct quote. A video is set like a TV news broadcast (with breaking news scrolling from right to left at the bottom of the screen) while the (You)Tuber says:

stay on our news we have a special guest his name is Chris Crocker also known as the Britney Spears' cry baby. Chris Crocker wants to tell us about his life and how

<sup>118</sup> *SuperMario* is a popular video game character and is a widespread 'meme' in *YouTube* videos.



he's supposed to succeed in his life Chris Crocker you're on.

Then the 'Best Video EVER!' appears with farting sounds in the background, followed by the respondent opening his mouth wide and saying:

oh sorry yeah sorry that's a very successful life you got there Chris I hope you make millions thank you.

Then follow the typed credits (accompanied by a heavy metal soundtrack):

thanks to our special guest chris crocker

The recontextualization of CC's initial video as a guest-appearance on a news broadcast functions both as criticism to CC and the insignificance of the initial video, and as criticism to the grounds of CC's celebrity status.

An embedding also corresponds the initial video; here CC's blinking follows the (You)Tuber's (Fig. 80). In embedding CC's video, the video expresses its response function, thus resuming the exchange (similarly to the 'exchange-narration' responses in the 'Where Do YouTube?' thread, cf. Chapter 5, Section 3.3.1).

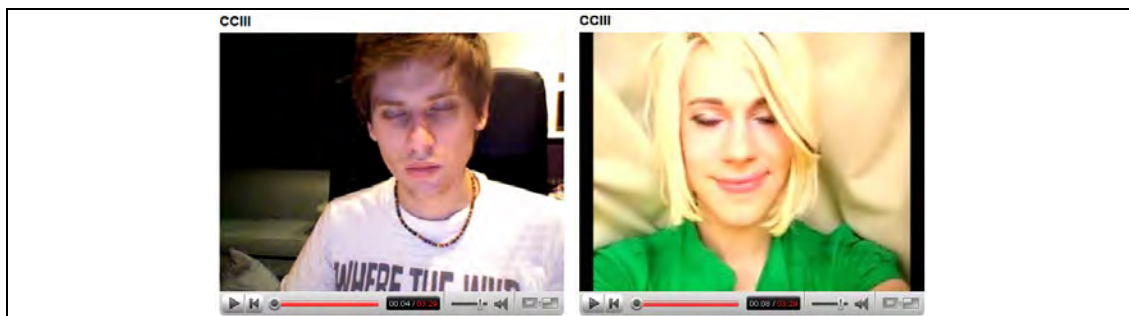


Fig. 80. Embedded initial video recapitulating the exchange.

A response (Fig. 81) represents a newspaper turning while increasing its dimensions (as in the '60s filmic conventions of news announcements). The newspaper masthead is 'The YouTube Times' and the front page features a black and white snapshot of the initial video with the title 'CRAZY QUEER INVADES YOUTUBE'. When the newspaper is at a close-shot distance, the embedded video plays. The response both celebrates CC's fame and evaluates CC's persona negatively (i.e., 'Crazy queer').



**Fig. 81.** The initial video embedded in a newspaper.

A response (Fig. 82) starts with typed titles scrolling from large to small from the bottom of the screen (i.e., recalling the opening of *Star Wars*, a popular ‘meme’ on *YouTube*). A superhero-masked (You)Tuber reads a newspaper titled ‘Chris Crocker’, with ‘To Reveal Big Secret TONIGHT’ as featured title. A ‘dialogue’ is constructed through reversed-angle shots (cf. Chapter 4, Section 1.3) of the superhero and CC (through shots of different videos of CC, including the initial one), in which the former tries to persuade CC not to reveal his secret identity. Since his various attempts do not succeed, the superhero blasts a bomb (a yellow balloon) and CC dies. Therefore, this response uses CC’s snapshots and CC’s utterances and recontextualizes them in a new dialogue.

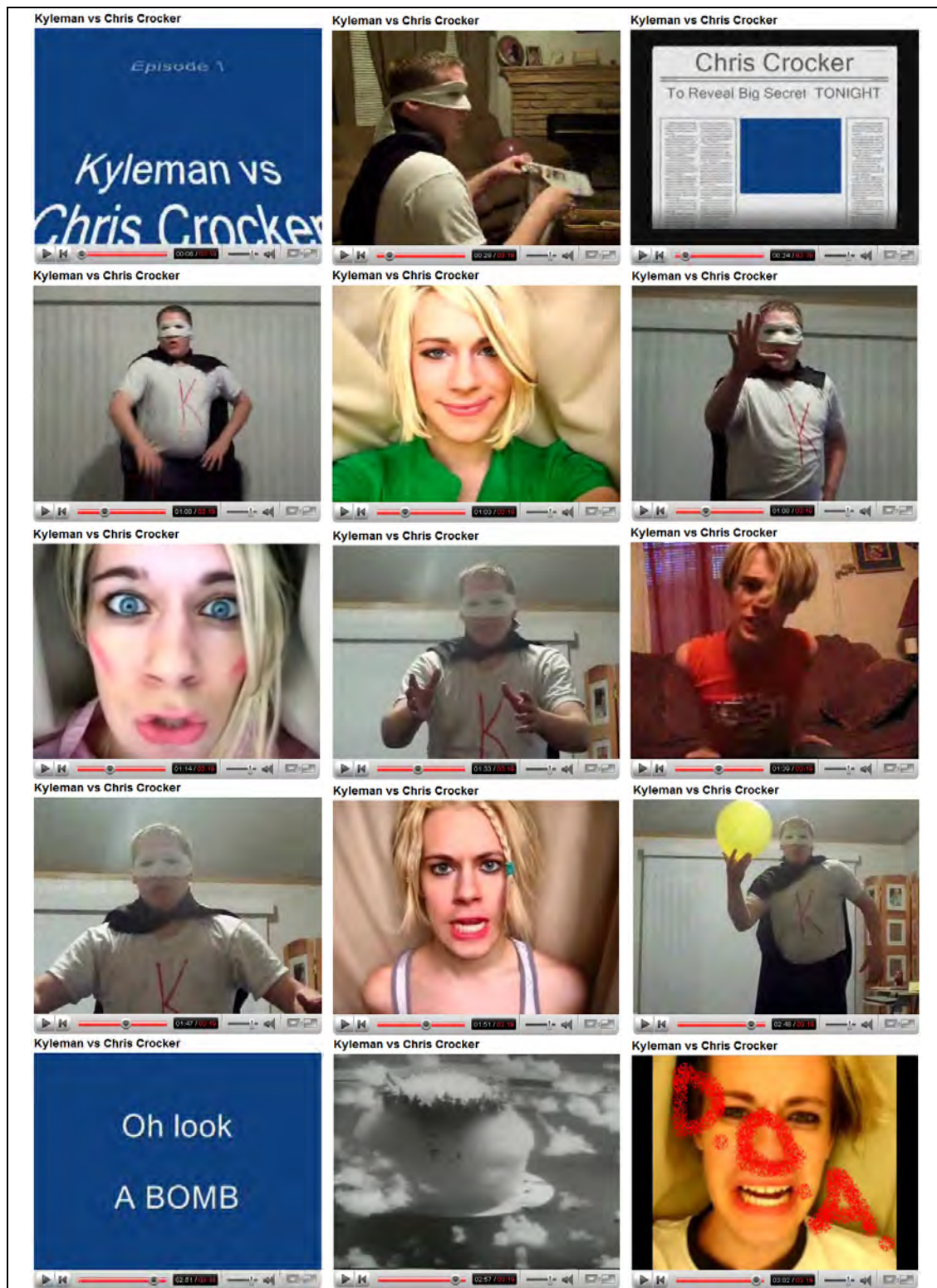


Fig. 82 Embedded material recontextualized in a dialogue.

### 3.3.2 Recontextualization of other CC's videos

Like the remixes of the initial video, also the responses which recontextualize other CC's videos can be (1) reposts, (2) more or less modified (and distorted or deformed) mashup videos, or (3) embedded material used as a quotation in the response. Many of them use – and transform – CC's most famous video 'Leave Britney Alone'.

#### a. *Reposting*

A response reposts a video of CC presented in the description as 'very rare':

This is Chris Crocker's new Music video to Britney's amazing Song "piece of me"  
EXTREMELY RARE

The practice of reposting famous videos is well established on *YouTube* (so that the search string 'leave britney alone' results in several hits of the original video reposted by different profiles). It is both a way of endorsing one's favourite video (and signalling it to viewers) and a means of increasing the views to one's channel.

#### b. *Editing*

Responses can remix various shots of CC's videos, thus creating mashups. Some of these are presented as a homage to CC, like a video titled 'Baby...you're the best', with a mashup of CC's videos on a unique soundtrack and some typing at the end:

congratulations chris for 100,000 subscribers and for entertaining us all

Other edited remixes clearly parody CC. One, titled 'Chris Crocker Is Pregnant!', represents a distorted image of a bare-chest CC, enlarged at the bottom so that the belly is expanded and recalls pregnancy (Fig. 83), accompanied by a celebrative (i.e., triumphal) soundtrack and a typed closing 'congratulations! Chris crocker'.



Fig. 83. Distorted ('pregnant') CC.

Some are ambiguous in their interpersonal meaning, such as a mashup of both CC's videos and of CC's appearances on TV shows, titled 'Chris Crocker when Opportunity knocks'. Another one, titled 'itschriscrocker in 30 seconds', is a

slideshow of snapshots of CC's videos, which typify CC's most characteristic pouting and 'outrageous' expressions. The snapshots are not modified, so that the mashup can be appreciated as honouring CC by CC's fans and, simultaneously, it can be enjoyed as a negative exposure of CC's 'silliness' by CC's detractors. Here again, it is the meaning-maker's interest which drives the specific interpretation of the video, independently from the uploader's intended meaning in posting the mashup.

Another remix, titled 'Chris Crocker Penis meal', uses CC's famous 'Leave Britney Alone' video with the voice sound manipulated and subtitled, so that now CC is crying for a relationship (Fig. 84), as the typed excerpt evidences:

all i want is relationship - then this new year... - no one bought because i was in it  
- any man... - one... - i stabbed him... - and there was a penis-meal.. - Bill! - i don't  
want to leave Bill...!!



**Fig. 84.** The rephrasing of the 'Leave Britney Alone' spoken text.

A response remixes CC's shots, while CC's words synch with the tune of the very famous 'Chocolate Rain' *YouTube* video (cf. Chapter 4, Section 3.1.2). By pairing and mixing the two *YouTube* celebrities' artefacts (also the 'Chocolate Rain' protagonist has appeared on TV shows, etc.), the video response plays with *YouTube* meta-references. This is an instance of a totally self-referential practice. Here again (cf. 3.2.1.a and 3.4.2.b), intertextuality plays a major role in the enjoyment of the video. In sum, in the elitist pleasure for implicit intertextuality, video-interaction is a typically post-modernist interactive practice.

### *c. Embedding*

Among the ones which embed CC's semiotic material, a video response (Fig. 85), titled 'Chris Crocker Frightens Me', recalls in its description the already cited meme 'reaction to...' (cf. 3.2.1): 'My reactions to a particularly disturbing "Chris Crocker Experience"'. The video features a boy's face looking at the screen (signalled by the screen light reflected on the boy's face); then the pc screen is filmed showing an excerpt of CC's video 'Eat My Corn Hole' while CC says 'kiss my ass and eat my corn hole'; then the (You)Tuber faces the camera with a worried and puzzled expression before raising his t-shirt onto his head.



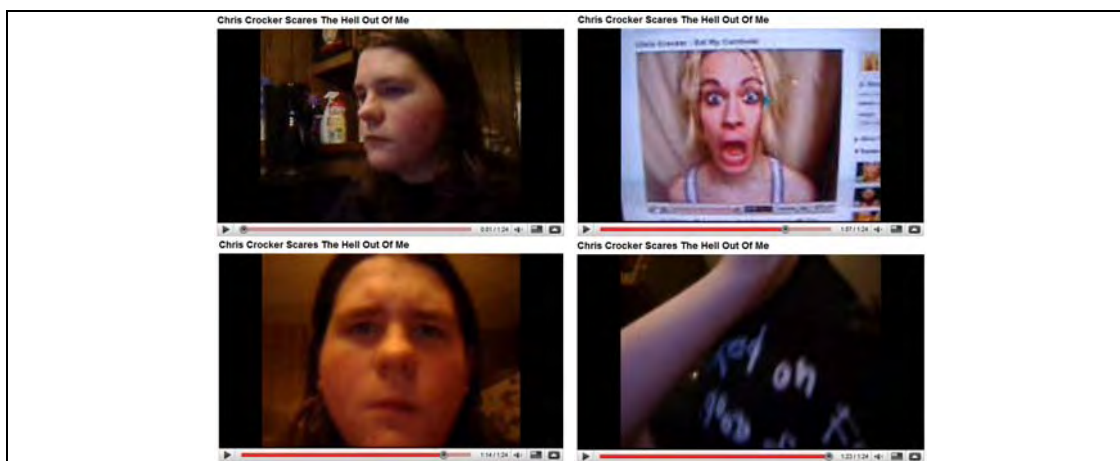


Fig. 85 A 'reaction to...' response.

Another quoting response features a masked (You)Tuber saying:

hello cc remember me? well if you don't remember me refresh your memory

followed by a mashup of some of the (You)Tuber's videos parodying CC. These are CC's videos, manipulated with a fast-forward voiceover – like the famous singing voice of *The Chipmunks* cartoon (another widespread *YouTube* 'meme'). The snapshots are accompanied by the typing 'I got the urge to take a really giant sh\*t in your mouth' (Fig. 86). Then the masked (You)Tuber appears and talks again – always speaking *Chipmunks*-like –, this time addressing CC's viewers:

It seems your acting again Chris Crocker. All those watching Chris Crocker it's time for you to **stop emulating your hero little loca** I know you learned a lot about how to tube on youtube Mr Crocker **let me advise you take your video camera throw it in a fire because you got no talent** Chris Crocker **do us and the whole world a favour** and get your fat ass out of youtube! rrrrr

The utterances marked in bold here above appear also typed overlaid on the screen. This time the embedded material is not CC's original, but the (You)Tuber's one. In other words, it is a self-quoting of the (You)Tuber's parodies of CC.

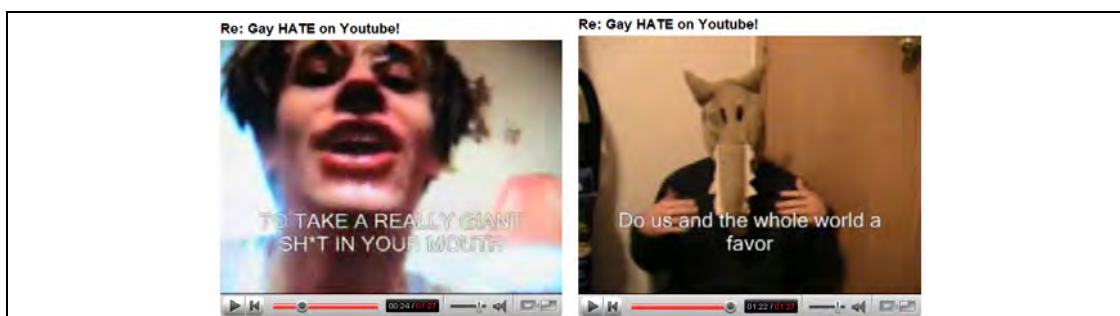


Fig. 86 Self-quoting response.

A response, titled 'Leave Chris Crocker Alone', features a mashup of CC's videos, other gay-related images, a typed 'fuck you to every gay-basher on youtube' and a boy singing his original text while playing on his guitar the music of the song 'Polly' by Nirvana. At the end, the (You)Tuber argues against gay-bashers (while – as often occurs – specifying that he is heterosexual). The text of the song addresses and argues against gay-bashers and CC's haters, as introduced in the video description, which quotes the definition of 'homophobia' and reports the song's text (another frequent practice for videos featuring original songs):

Except when clearly expressed, as in the above discussed cases, it is often difficult to determine whether these remixes are intended to insult or to praise CC. Exemplary of the extreme polyvalence of remixes is a response featuring a man rapping his original song, interposed by another man wearing a blond wig who sings more melodic pieces of the song, and also by CC's snapshots variously modified. The text of the song is a quite insulting mockery of CC, as the following excerpt illustrates:

Don't Cry For Britney, Chris Crocker  
She doesn't even know you exist  
And if she was in bed with you  
You wouldn't know what to do.  
AND HIS FACE IS A MESS!  
Chris Crocker Chris Crocker  
Doesn't like verginer...wears eye liner!  
Chris Crocker Chris Crocker  
Dances and sings...wears women's underthings  
Chris Crocker Chris Crocker  
Wanna be sedated...he took a dump and ate it!  
you's a freak but you aint unda cova  
u make love to yo own brother!  
Chris Crocker is you outcho mind?  
Sticking things all up in yo' behind?  
AND HIS FACE IS A MESS  
AND HIS FAVORITE FLAVOR IS COCK

Before reporting the words of the song, the video description specifies its author's positive attitude towards CC:

NOTE: LOVE IT OR HATE IT, THIS IS COMEDY. I DO NOT HATE CHRIS CROCKER, I think he's a talented young man so much so that I created this video. Accept it for what it is, the same way your parents try and accept you even though you disappoint them at every turn. (that too was comedy, you illiterate fucks).

In other words, parodying, mocking or even insulting coarsely can be meant as a homage addressed to the parodied (You)Tuber, i.e., an act of thanking the targeted character for providing (You)Tubers with precious material which prompts their (spoofing) creations.

Indeed, many remixes are simply playful and creative transformations of previous texts and are posted as such, sometimes promoting the creation itself or part of it. For

example, a response is a repeated two-image slideshow of CC crying (and the callout ‘leave britney alone!’) and of a bald Britney Spears. The soundtrack is an original song titled ‘Cry Chris Croker’, while a typing, appearing here and there during the video, invites to download the tune (Fig. 87). The video description gives the link to *iTunes* where the song is downloadable. In other words, the response promotes the auditory material which was created for it.



Fig. 87. A self-promoting remix.

As Willet (forthcoming) rightly argues, it is often the case that parodies on *YouTube* are not explicit as to whether they have to be intended as homages or criticisms. In many cases, it is the pleasure of spoofing in itself which drives the creation, while the (You)Tuber is not interested in communicating explicitly her attitude towards the parodied. Again, it is not the sign-makers’ intended meaning which matters, but rather the pleasure of a creative transformation of other people’s texts, which finds its counterpart in the viewer’s enjoyment in understanding the intertextuality (i.e., the parodied source). When the response does not parody the initial video but other CC’s videos, the enjoyment is given by the retrieval of the implicit reference (i.e., not signalled by the response link to its original parodied source), which requires knowledge of the parodied’s previous semiotic acts, within and outside *YouTube*, as in the case of a remix which embeds CC’s TV appearances.

### 3.4 Original spoofs

A different type of ‘spoofing’<sup>119</sup> responses (34) uses original material – instead of snapshots of CC’s videos. These responses typically feature a (You)Tuber parodying *Chris Crocker* through selected salient elements of CC’s character, e.g., a blond wig, the make up, the blanketed background, CC’s characteristic verbal rhetoric (the so-called ‘bitch-speech’). Rather than by virtue of original quotations, these responses establish cohesion by means of ties of repetition (either of the initial video or of salient elements of its character and previous deeds).

<sup>119</sup> Following Willet (forthcoming) I am here using ‘parody’ and ‘spoof’ interchangeably.



### 3.4.1 Spoofs of the initial video

As anticipated in the related section, some of these original spoofs are also corresponding responses (i.e., featuring the (You)Tuber wearing a blond wig and blinking, thus simultaneously spoofing CC and corresponding the initial video); one of these (Fig. 88), titled ‘the Best video ever 2’, uses drawing instead of the shot of the (You)Tuber, so that a stylized bald<sup>120</sup> man in an emerald-green blouse looks at the camera, raises his hands and blinks, while a ‘suspense’ soundtrack (i.e., drums increasing their rhythm) bursts into a scream of people exulting.



Fig. 88. A corresponding original spoof made by means of drawing.

### 3.4.2 Spoofs of CC's character

Instead of corresponding the initial video, other responses take up salient elements of the character as signifiers of CC, to enact an original parody.

#### a. *CC's imitation as a humorous performance*

Some original spoofs are clearly intended to produce humour and get the viewers enjoy the video creation; for example, in a response (Fig. 89), the camera shoots a pair of feet walking indoor while the voice of a boy says

tatantatatan I'm talking to you because I'm much prettier than Chrissy, you got some competition

When the filmed feet arrive in the bathroom, the camera moves up and shoots the mirror, which reflects a boy wearing a blond wig. The spoof of CC is taken as a chance of enacting a suspense plot with a humorous revelation at the end.

---

<sup>120</sup> The baldness may refer to the famous Britney's bald head or to CC and the title (i.e., the 'very' best video ever would feature a bald CC).



Fig. 89. A humorous plot parodying CC.

b. *The spoofed character intertextually contextualized*

Responses can combine the spoof with widely known *YouTube* memes, by inserting the spoofed character in another popular *YouTube* genre, thus producing a highly enjoyable intertextual reference. One refers, to the already cited (cf. 3.3) ‘Response to 2 girls 1 cup’ format, with the title ‘2 girls 1 cup reaction from Chris Crocker’. The video is introduced with the following typing on screen:

O.K. I was dumb enough to find out what the "2 girls 1 cup", video was all about. I wish I hadn't, but it's too late now  
Well, when discussing script ideas, one of us idiots thought of this one, but not with the idea of dissing chris crocker.  
That being said we hope you enjoy  
CHRIS CHROCKERS REACTION TO:  
"2 GIRLS 1 CUP"

Thus, the introduction recounts the authors’ design phase of their response to CC’s initial video and their idea of combining the response with a well-known meme on *YouTube* videos, i.e., the enacting of a viewer’s reaction while watching the ‘2 girls 1 cup’ porn film. Then the video features a boy wearing a blond wig and make up (thus personifying CC), with a whitish blanket (a duvet, actually) behind him, looking at a pc screen and saying that he is going ‘to check up 2 girls 1 cup’. He watches the screen (you can hear noises of sexual intercourse in the background), crying ‘oh my god!’ and eating chocolate ice cream (putting it in and out of his mouth, with a clear sexual reference); then he says ‘oh give me a close up’, while laughing with all his mouth smeared of ice cream. Fig. 90 represents the salient snapshots of the response.



Fig. 90. The ‘2 girls 1 cup reaction’ intertextual original spoof.

The reaction (which could well be posted also as a response to other ‘2 girls 1 cup’ videos) combines CC and the ‘2 girls 1 cup’ meme, by implying that CC’s sexual preferences are so ‘pervert’ to ask for a ‘close shot’ (rather than to express disgust as reactions to ‘2 girls 1 cup’ usually represent). The introduction however hedges the interpersonal meaning of the video (‘not with the idea of dissing chris crocker’). Again, parodies and enactments are made for the sheer pleasure of it, rather (or more) than to criticize or ‘bash’ the parodied character.

A further response uses the ‘blender’ meme, i.e., an animation video featuring a blender where objects, animals or even humans are blended until they are splashed and smashed. The response in the thread, titled ‘chris crocker gets pwned’ (Fig. 91), features a drawn *Chris Crocker* in a blender weeping the ‘Leave Britney Alone’ refrain. The blender starts blending until CC is smashed; a dog gets out of the blender and screams ‘freedom’; then the typing ‘freedom to blend’ appears.

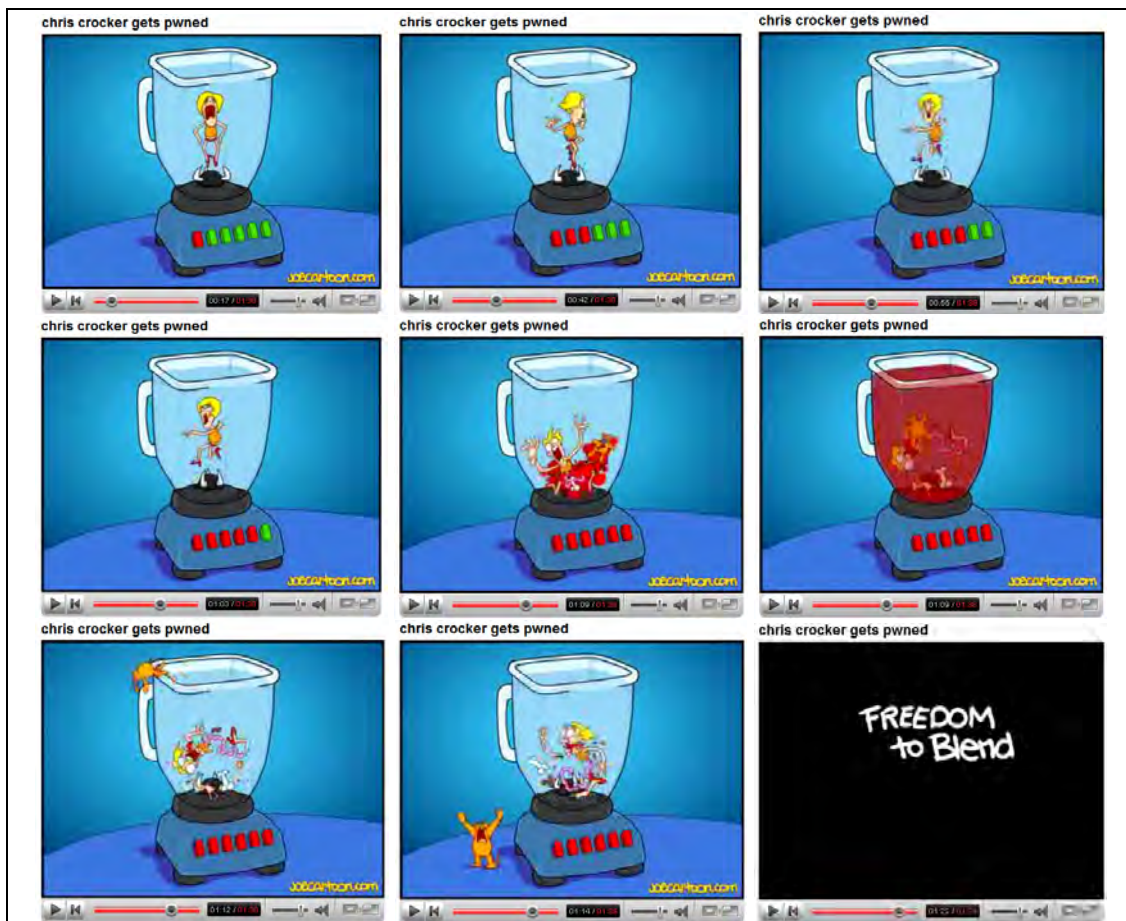


Fig. 91. the blender intertextual original spoof.

Here again, independently from the author’s intention of literally ‘bashing’ CC, the real enjoyment, both in creating and in watching the video, resides in understanding the mixed intertextuality; it is really the pleasure of playing with genres and

combining them, of taking up other people's signs and transform them in new texts which is sought for here.

These are further cases (cf. also 3.2.1.a and 3.3.2.b) where the implicit reference to another text makes the video even more enjoyable to experienced (You)Tubers; it appeals to an elitist feeling and gives the viewer satisfaction for having grasped the reference to a fact which is accessible only to insiders.

What all these videos with intertextual references have in common is that they could be posted as well as a response to their intertextual reference source; in this case the response meaning would be recontextualized in a new exchange and the type of relatedness would change. This often occurs, when an initial video is removed or when it gets old and the respondent wishes to change its response link to a more recent or popular one. As one sometimes reorganizes the files and folders on a pc, (You)Tubers often reorganize the links of their videos to make new meanings out of these relations, according to their renewed interests. Even more, the 'hypertextual' nature of the interface allows a (You)Tuber to exploit all available options to have a video linked to others, and any link does not exclude another. So, for example, by virtue of its title, the CC's reaction to '2 girls 1 cup' appears both in the related sections of other 'reaction' videos and a response to CC's 'Best Video EVER!').

c. *Spoofs of CC's previous videos*

Like the remixes, also some original spoofs parody the 'Leave Britney Alone' video (rather than the initial video). A response, titled 'Leave the "Leave Britney Alone Guy" Alone!!', features a man wearing a blond wig, with a blanket in the background. His speech, pace and sobbing rephrase exactly CC's, while changing some of the words, as evidenced here in the parallel transcription of the two videos:

**Leave the "Leave Britney Alone Guy" Alone!!**

And how fucking dare any of you make fun of the Leave Britney Alone guy after all he's done for you. [sobs]

All you people care about is readers and money [sobs]. He's a HUMAN. She is a HUMAN. Whatever. [sobs] What you don't realize is that he's making new millions and millions of dollars out of YouTube so now and all you can think to do is make fun of him in videos. Leave him alone. He hasn't made a video in days. His video was called Leave Britney Alone for a reason because all you people want to do is make fun of Britney Britney Britney Britney Britney. [sobs] Leave him alone. [sobs] You're lucky that he

**Leave Britney Alone (CC's original)**

And how fucking there anyone out there make fun of Britney after all she's been through. [sobs] She lost her hair she went through a divorce. She had two fucking kids her husband turned out to be a user a cheater and now she's going through a custody battle. All you people care about is [sobs] readers and making money out of her. She's a HUMAN! [sobs] What you don't realize is that Britney is making you all this money and all you do is write a bunch of crap about her.

She hasn't performed on stage in years. Her song is called gimme more for a reason because all you people want is more more more more more. [sobs] Leave her alone! [sobs] You're lucky she even

even made a YouTube video for you BASTARDS. [sobs] Leave the Leave Britney Alone guy alone. [sobs] please. [sobs] Paris Hilton talked about professionalism and said that Britney Spears would have done it if she was a professional. She would have done it no matter what. Well speaking of professionalism how is that fucking Paris Hilton gets a television show it sucks and it's not funny, cancel the show right now and leave the Leave Britney Alone guy alone. [sobs] Leave him alone. [sobs] If you have a problem with the Leave Britney Alone guy you deal with me because he's not sane right now. [sobs] Leave him alone.

performed for you BASTARDS. [sobs] Leave Britney alone! [sobs] Please. [sobs] Paris Hilton has talked about professionalism and said if Britney was a professional she would have pulled her off no matter what. Speaking of professionalism when is professionalism to publicly bash her when she's going through a hard time. [sobs] Leave Britney alone! [sobs] Please. [sobs] Leave Britney Spears alone right NOW! I mean it. Anyone who has a problem with her you deal with me because she's not well right now. [sobs] Leave her alone.

The video concludes with some 'director's cuts', i.e., rehearsal mis-scenes (e.g., the blanket falls on his head; he bursts into laughter while shooting); then the boy appears without the wig and says, in his own intonation, 'thanks for watching, don't forget to subscribe' (Fig. 92 represents some salient snapshots of the video).



Fig. 92. A 'Leave Britney Alone' original spoof.

Here the (You)Tuber enacts an *ad litteram* parody of CC's 'Leave Britney Alone'. He follows quite precisely the original text, while introducing some significant variations. One deals with Paris Hilton's unmerited TV show, not mentioned in CC's original video (thus communicating his own opinion on the matter by introducing it in the parody). The other – more crucial in setting the sense of the parody – is a replacement of participants: he substitutes CC with himself – imitating CC – as the represented participant, and substitutes Britney Spears with the 'Leave Britney Alone guy' (i.e., *Chris Crocker*, but without ever using the name) as the topic of the spoken text (i.e., the spoken protagonist). Through this double substitution, the spoof achieves an additional sarcastic meaning, in consideration of the often cited fact that CC became famous thanks to CC's defence of Britney Spears (CC's detractors restate this fact as CC exploiting Britney Spears and her misadventures so as to achieve visibility). This parody can be then interpreted as the (You)Tuber exploiting CC's achieved fame – now fallen down – so as to gain visibility. In this sense, it is also an implicit counter-response to the 'Best Video EVER!'. Indeed the initial video description states the 'point' of the video as proving that CC is still famous. In turn, the parody asserts that CC is going through a bad moment in fame as well, by affirming that CC 'is not sane' and 'hasn't made a video in days', and by metaphorically relating (i.e., substituting) CC to Britney Spears in the context of the



‘Leave Britney Alone’ video (i.e., of a guy defending Britney Spears while she is going through a bad moment in fame). Parodies (and remixes) are impressively numerous as responses to the ‘Leave Britney Alone’ video. Yet the reach of the video which made CC famous is so wide that their parodies are found also among the responses to the ‘Best video EVER!’ video, posted several months later.

d. *The ‘Britney Spears’ element as the prompt of the spoof*

Also the ‘Britney Spears’ element connected to CC’s history and character can be taken up in the responses. For example, a (You)Tuber, wearing a blond wig and a big fake breast (i.e., two balloons under the t-shirt), says ‘it’s Britt Knee, bitch’ (the text is also subtitled). Then he dances while lip synching a song by Britney Spears, with the modified title ‘Gimme Morse’ (instead of ‘Gimme More’), and a photo of Mr Morse appearing when the song sings ‘More’ (Fig. 93).



Fig. 93. An original spoof taking up the ‘Britney Spears’ element.

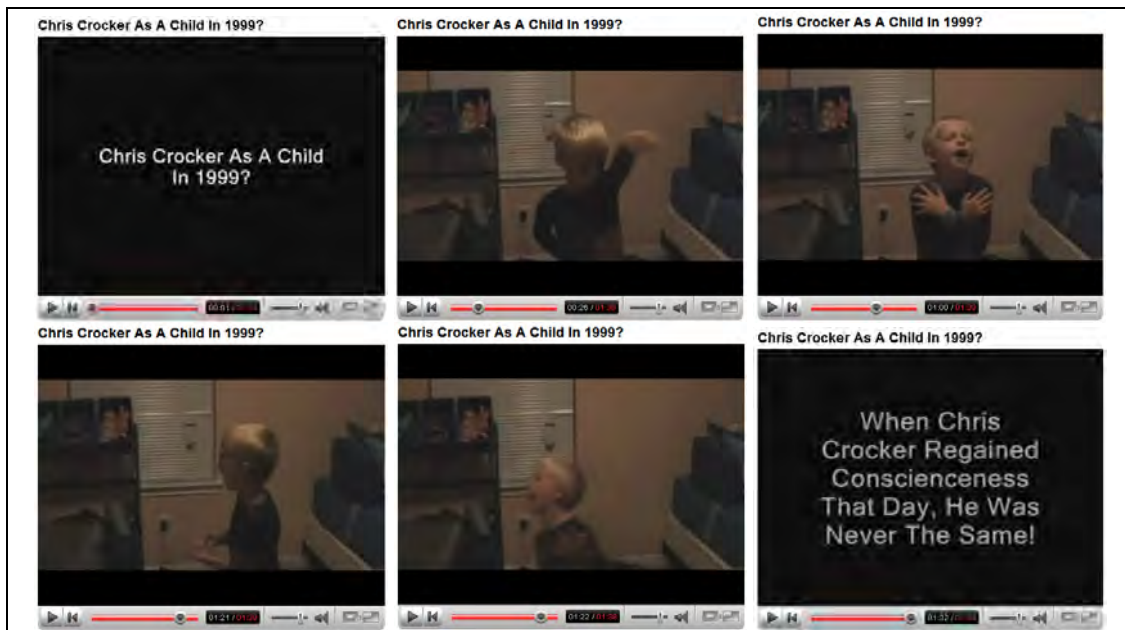


Fig. 94. A ‘plot-enactment’ original spoof.

Britney Spears is often used as a reference also by means of the soundtrack playing

her songs. So, the response shown in Fig. 94 opens with a typed ‘Chris Crocker As A Child In 1999?’. A child dances while Britney Spears’ song plays ‘hit me baby one more time’. He feigns to fall, struck by something (like a heart attack), then it appears the following typing: ‘When Chris Crocker regained conscience that day he was never the same’. This response makes a mockery of CC’s declared fanaticism for Britney Spears by enacting the origin of it and, by inference, of CC’s ‘insanity’.

In sum, original spoofs can parody either or both the initial video (and thus also correspond it) and/or CC, by taking up a salient element of CC’s persona. They often parody CC’s most famous video (Leave Britney Alone) and sometimes spoof CC’s idol, Britney Spears. Like the remixes, also the original spoofs can be more or less explicit as to whether the parody is hostile to or celebrative of the parodied character. Some are more clearly emulations or imitations of CC’s style (for the difference between imitation and parody, cf. Willett, forthcoming). At a general level, parodies are surely intended to entertain humorously the viewer. To do so, given the large number of them which target the initial video and its author, they strive to produce variation, and the parodied character is taken as the chance to represent an original performance and to communicate one’s own distinctiveness.

### 3.5 Inferential responses

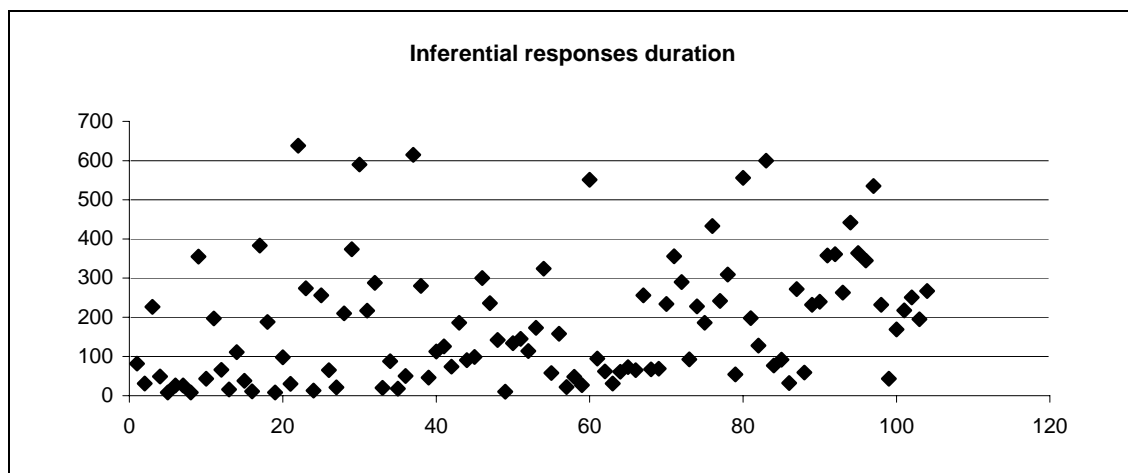


Fig. 95 Distribution of the inferential responses according to their duration: 104 videos; mean 182”.

Unlike the typologies discussed so far, 1 out of 6 responses (104 videos; cf. Fig. 95) does not explicitly refer to the initial video. These responses neither deploy its multimodal pattern nor address CC directly (by naming CC, or using snapshots of CC’s video) or indirectly (by metonymic imitation, e.g., by using a blond wig). Instead, they take up a backgrounded element of the video or of its character (and history) and make it their salient prompt to which to respond. These are here called ‘inferential’ responses because they need a greater deal of inferential work for their relatedness to be grasped than the other responses just analysed.

### 3.5.1 Inferential prompts from the initial video

#### e. *St. Patrick's day-related responses*

Among the ones which take up a background element and make it salient (thus actualizing it as a responded prompt) are two responses whose topic is St. Patrick's day (Fig. 96). One is a videoblog of a guy wishing a happy St. Patrick's day to viewers, wearing an emerald-green t-shirt and indicating it by means of verbal language and by pointing it to the camera. The other, is made of a *Second Life* animation representing '3 friends' who 'go out on st. patricks day and get drunk', as the video description explains (the pub has emerald-green lamps, stools and sofas).

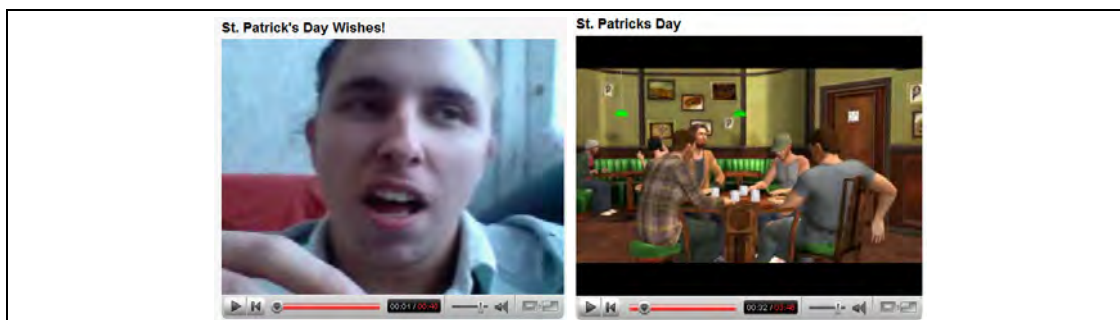


Fig. 96. St. Patrick's day-related responses.

These responses would seem totally unrelated to the initial video, unless one considers that the latter was posted on 17 March, i.e., St. Patrick's day, and that CC's blouse has the emerald-green colour nuance, which is traditionally associated to St. Patrick's day celebrations. In other words, the responses interpret a background element of the initial video according to the respondents' interest and make it salient by turning it into the responses' topic. It may be considered an instance of topic diversion. However the case is more complex. Indeed, in order to grasp the relatedness of the exchange, viewers need to know (1) the conventional semiotics which surrounds the occurrence of 17 March in the Irish world and in their immigration areas, such as U.S. – i.e., St. Patrick's day, whose celebrative colour is emerald-green, which is used on clothes worn on that day –, (2) link it to the day of posting of the initial video and (3) to the colour of CC's blouse. Even better, viewers need to be interested enough in searching for clues of relatedness (i.e., retrieving all this knowledge needed to perceive the relatedness), otherwise they will discard the response as totally unrelated (also by virtue of the fact that many responses in the thread are indeed unrelated, as discussed in Section 3.9).

#### f. *Make-up-related responses*

Analogously, two responses take up the 'make up' element of the initial video, so that a respondent looks at the camera and shows how to create the 'Perfect Pout' (the title of the video; cf. Fig. 97), by applying an exaggerated makeup herself. This



response may be interpreted as totally unrelated; however, if one recalls the response of the girl praising CC for ‘the best make up that I have ever seen done’(the commentary discussed in Section 3.2.1.b), the ‘make up’ element of the initial video may be considered as the prompt actualized in both responses, one commenting CC’s make up, the other showing didactically (and ironically) an alternative one.



Fig. 97. The make-up related response.

g. *Paratextual elements-related responses*

Finally, a response takes up the temporary description of the initial video about AIDS (cf. footnote 108), and features a slideshow of images and typing focusing on the disease.

### 3.5.2 Inferential prompts from the character

a. *Homosexuality as ideational topic: the thread as a forum discussion*

Nine video responses take up an element of the character, i.e., CC’s declared homosexuality (without ever mentioning CC) and make it their topic (either supporting gay issues or ‘bashing’ them), so that the initial video and its notorious character are used as the chance to set up a forum of discussion on gay-issues.



Fig. 98. A gay-hate inferential response.

Among the gay-hate responses, one (Fig. 98) features an elderly man speaking about the ‘God of the Bible’ who ‘hates fags’, interposed by a slideshow of images

representing boards, posters and photos of protests against gays. The response is further responded by a number of videos arguing against his positions (cf. Section 4), so that a forum of discussion is generated and extends to other levels in the thread.

Other responses discuss the gay issue in more personal terms, e.g., a boy presents himself as gay, videoblogs about his troubled relationships and asks for advice, while another boy videoblogs on his outing to parents and friends.

Responses can relate to the gay-issue very tightly (as in the above examples) or quite loosely; in this latter case, they refer to it implicitly, as included within the realm of 'discriminations'. For example, a video titled 'Ode to Adrian Piper' shows, the following statement set up in the form of a letter, typed in white on a black screen, where blanks can be filled by any type of discriminated difference (i.e., gender, race, sexual orientation, religion and beliefs etc.):

Dear Friend, I am \_\_\_\_\_. You may not have realized it when you made/laughed at/ agreed with that \_\_\_\_\_ comment. I regret that my \_\_\_\_\_ makes you uncomfortable as much as that your \_\_\_\_\_ attitudes alienate me. \_\_\_\_\_ people are not always visible. We witness hatefull comments towards \_\_\_\_\_ people in public spaces everyday. This often makes us feel like disappearing, that we are unacceptable in the eyes of other people. It manifests itself in high suicide rates, substance abuse, anger, pain and in art.

Adrian Piper is a conceptualist artist whose work focuses on issues dealing with race and gender discrimination. By linking this video as a response to CC's, the text recontextualizes itself and the blanks can be more promptly filled with gay-related lexemes (rather than, e.g., with race-related ones). Arguably, the intent of the respondent is to have the video gain visibility by linking it as a response to a very popular one. The selection of CC's video among the several (thousands of) popular ones on *YouTube* has clearly been prompted by the character's declared sexual orientation and by CC's frequent videoblogging on gay-related issues.

*b. Homosexuality as interpersonal prompt: the thread as a meeting place*

A much larger set of these responses (39) does not make homosexuality a topic but rather uses it interpersonally, so that the thread is used as a sort of a meeting place (rather than a forum discussion) for gay-oriented or transsexual participants. Hence respondents, whose characters deploy elements of gay or drag queen aesthetics (Fig. 99.a) simply videoblog their daily experiences (e.g., a boy videoblogs on his life and on a former – heart breaking – love-relationship with another boy), or dance chest-bared and pout in front of the camera (Fig. 99.b). Three responses (responded in their turn) give way to a lesbian meeting place and forum (e.g. Fig. 99.c).

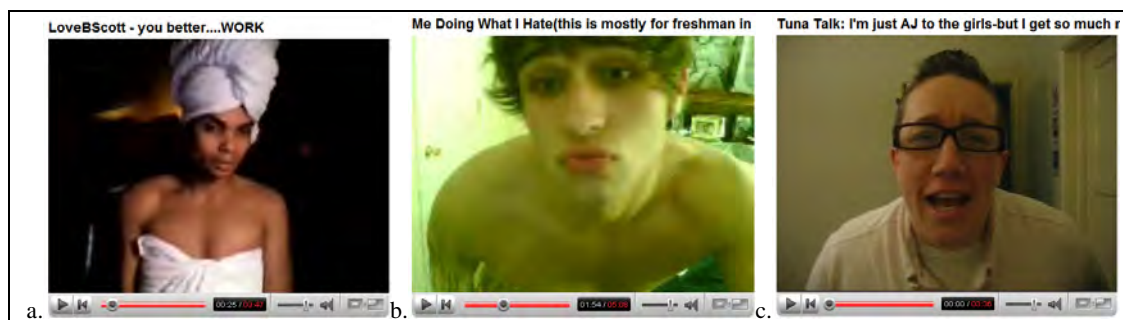


Fig. 99. Inferential responses which use the thread as a non-heterosexual meeting place.

Without ever referring to it, these responses take up interpersonally the ‘non-heterosexuality’ element at large (from homosexuality, to drag queens, cross-dressing and transgenderism) which connotes CC’s character. In other terms, these respondents post their videos as a response to a very popular one which is likely to be viewed (also) by non-heterosexual viewers. The fact that CC’s videos are a basin – a meeting place – for non-heterosexuals is further confirmed by a response, titled ‘A LITTLE TOUR OF MOLLYWOOD - Toronto’ (Fig. 100), which advertises gay-places in Toronto and promotes the town as ‘a great place for gays’.



Fig. 100. Toronto advertised as an ideal place for gays.

The ‘meeting place’ interpretation of the thread by this type of responses is also confirmed by the fact that many of the respondents post more than one response to the initial video, in various periods of time, often videoblogging on different topics. So, for example, a blond-haired (You)Tuber posts four different responses, featured with a whitish blanket in the background. In one (Fig. 101.a), she<sup>121</sup> videoblogs about her hard times in being a transgender and against spoofers and haters, while in another (Fig. 101.b), she shows how to make bourbon and coke. Later in the thread, with a different background, she also posts a blinking corresponding response titled ‘THE MOST BEAUTIFUL VIDEO EVER!’ (Fig. 101.c).

<sup>121</sup> ‘She’ is the pronoun the videoblogger uses when addressing to herself.



Fig. 101. Multiple posting in the thread used as a meeting place.

In another example, a (You)Tuber – who declares his homosexuality on his channel – videoblogs once about his new hair colour and asks his viewers their opinion on it (Fig. 102.a.); later, he posts another videoblog on his new blond hair and asks again the viewers' opinion (Fig. 102.b.); while earlier he had videoblogged about his new propositions for the new year (Fig. 102.c.).

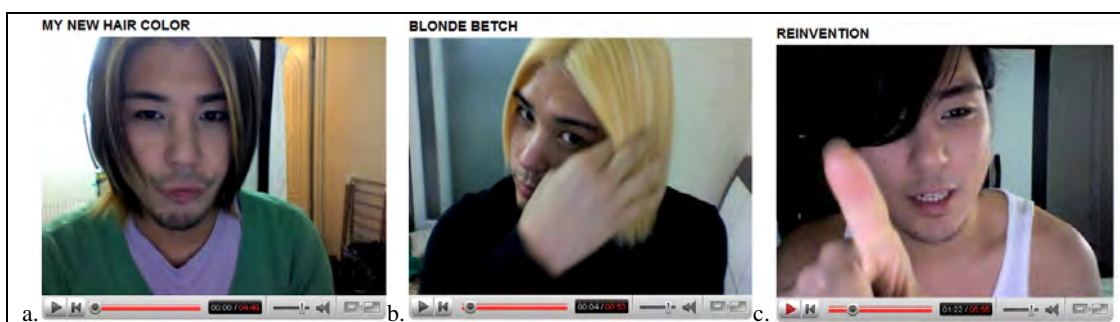


Fig. 102. Multiple postings (the last screenshot on the right: 'delete' simulating gesture).

Occasionally, these videobloggers may also mention CC, like the one just discussed, who, talking about reinventing himself (for the forthcoming New Year) announces that he has cleaned out his videos and 'all that hips about chris crocker and just delete delete delete', accompanying the spoken 'delete' with the index finger gesture which simulates the pushing of a button on the keyboard (Fig. 102.c).

These videoblogs often imitate (or even emulate) CC's aesthetics, in their intonation and in their 'bitch-speech'. One implicitly acknowledges the likeness, when, after lip synching, pouting and wearing lipstick, says 'hello I'm not Chris Crocker's boyfriend I'm a sexy rock star'. Of course, these aesthetic features are not just CC's, and are rather signifiers shared by a semiotic space which refers to the drag queen movement. Therefore, when there is no explicit reference to CC, it is often hard to tell whether these (You)Tubers emulate CC or if they simply share the community-specific aesthetics. What is certain is that they identify a famous personage of this movement to whom they link their semiotic acts so as to reach a specific audience.

c. *Sexual explicitness-related responses*

Instead of the ‘non-heterosexuality’ element, other responses (21) take up the ‘sexual explicitness’ element which characterizes CC’s persona in many videos and thus post mildly sex-related videos (e.g., featuring two girls, one slapping the other’s bottom) or whose titles promise porn content, which is however not fulfilled in the video. Often this misleading porn title is a tactics for gaining views, while the video features a couple of shots of girls pouting in front of the camera followed by Rick Astley singing his once famous song ‘Never gonna let you down’, used as a substitute for taboo-content on *YouTube*<sup>122</sup>. Indeed, due to the Website content prohibitions (cf. Chapter 4, Section 1.9), porn videos are normally removed from *YouTube*.

d. *Britney Spears-related responses*

Some other videos (15) take up an element of CC’s (You)Tubing history, thus they are topic-related to Britney Spears, who is CC’s declared idol and whose cried defence made CC famous in the ‘Leave Britney Alone’ video. Hence, the response may be the posting of one of the singer’s music video clips or a TV excerpt dealing with her. The ‘Britney Spears’ element is taken up here instantiating a different type of prompt-response relation than in the commentaries or in the parodies (cf. Section 3.2.2.b and Section 3.4.2.d). Indeed, here the singer’s video clips are (re)posted as they are<sup>123</sup>, with no explicit reference to CC nor intent of producing a parody. Simply, CC’s fanaticism for Britney Spears has prompted the (You)Tuber’s interested posting of a semiotic act related to the singer as a response to the initial video, maybe assuming CC’s audience includes Britney Spears’ fans. Also here, as in all these types of responses, a lot of inferential work is needed in order to grasp the relatedness in the exchange, if the viewer is interested in doing it.

e. *Fame-related responses*

Three videos take up the ‘fame’ element of CC’s character; two are posted by the same (You)Tuber and are filmed with the characteristic modality of paparazzi shootings (i.e., outdoor, crowded and noisy environment, camera-flash noises, no firm focus, the camera moving forward and shooting a character wearing a hood on the head so as to conceal the face, who often goes out of focus; the camera following the character from behind while the latter increases the pace, etc.; cf. Fig. 103).

---

<sup>122</sup> Rick Astley’s music video clip is widely used in this type of videos to disappoint viewers’ expectations on porn content; this practice is known as ‘Rickrolling’, as comments to these videos often admit, e.g., ‘oh my god! this is the third time that i got rickrolled this week!!’.

<sup>123</sup> Many of these videos are later removed by the Website owners due to copyright violation. Copyright is also the reason why no screenshots of these videos are presented here.





Fig. 103. Paparazzi-like inferential responses.

One of these responses' description is 'Celebrities shots', the other is 'Followed by paparazzo in West Hollywood'. Also these responses take up an element of CC's character, i.e., CC's celebrity, as an occasional prompt (and without ever referring to CC) to post their performance/enactment and link it as a response to the initial video.

---

In sum, these responses relate inferentially their content to some elements of the initial video (i.e., the St Patrick's day of its posting or the AIDS issue of its temporary description) or of its character (i.e., its sexual orientation, its sexual explicitness, its Britney Spears fanaticism or its notoriousness). CC's sexual orientation can be taken as an ideational prompt, so that it gives way to a debate on gay-related issues, or it can be taken as an interpersonal prompt, so that the thread is used as a meeting place for non-heterosexual participants.

The response's relation with the initial video is totally implicit, so that it needs to be established by the viewer by means of inferential processes. Even better, it may not really matter whether the viewer does indeed establish a relatedness between the response and the initial video. What seems certain is that the respondent, in deciding how to get the best from the use of the response link to her video, has selected CC's 'Best Video EVER!' on the basis of some criterial aspects of the video or of its character which could suit her purposes. Here again cooperation, relevance principles and notions of coherence are not sought for, except by viewers who – as Halliday rightly argues – keep insisting in finding clues for coherence (cf. Chapter 2, Section 2). Yet, they *can* keep insisting *whenever* they are interested in doing so.

Clearly, in some cases, as in the responses which use the thread as a meeting place for a certain typology of viewers, the initial video has not prompted the respondent to *make* a video response. Rather, the initial video has been sought for, in view of finding an apt video to which to link one's own (already created) as a response to it. In the St. Patrick's Day example, it could be both cases, i.e., that the view of CC wearing an emerald-green blouse in a video uploaded on St. Patrick's Day has prompted the respondent to respond to it by wishing a good St. Patrick's Day, or, in turn, that the respondent wanted to post a video wishing a good St. Patrick's Day and

has selected CC's as suitable to his interests (i.e., to create a response link).

Whether the initial video prompts the creation of a video response or it is just an apt chance used to establish a response link to an already created video, the underlying prompting-responding mechanism is by no means different. Indeed, it is the view of a video which triggers the creation of an interested relation, so that it prompts to link a video as a response to it according to the respondent's purposes.

The type of relatedness established here is the mere fruit of inferential processes, so that the viewer – the researcher, in this case – has sought for the coherence between two texts which had no explicit reference one to another. Indeed as Halliday rightly affirms (cf. Chapter 2, Section 2):

Texture is a matter of degree. It is almost impossible to construct a verbal sequence which has no texture at all – but this, in turn, is largely because we insist on interpreting any passage as text if there is the remotest possibility of doing so. We assume, in other words, that this is what language is for (1976: 23)

In other terms, Grice's cooperative principle can be said to be at work in interpreting an act which is formally connected to another (by means, in our case, of a response link on the interface established by the (You)Tuber). However, it is evident here – e.g., in the case of a video featuring a videoblogger with a gay aesthetics dancing and pouting in front of the camera posted as a response to the initial video – that the ways in which the relatedness has been established by the interactant is a different one, a very loose one. It follows more individualized (and less cooperative) interests, in view of exploiting whatever element the initial video prompts, which can establish even the remotest relatedness link to another, according to the interactant's interests. Analogously, in the viewers' perspective, if they are interested in retrieving any possible relatedness link between the two acts (as the writer did for the purposes of the present work), they will activate all types of inferences so as to find any clue for coherence in the exchange. In turn, if their interest is different – e.g., to search for responses which, say, criticize CC – they will not bother to activate any 'remote' inferential process and rather discard these responses as unrelated, or, rather, as irrelevant for their purposes.

### 3.6 Responses with secondary reference to CC

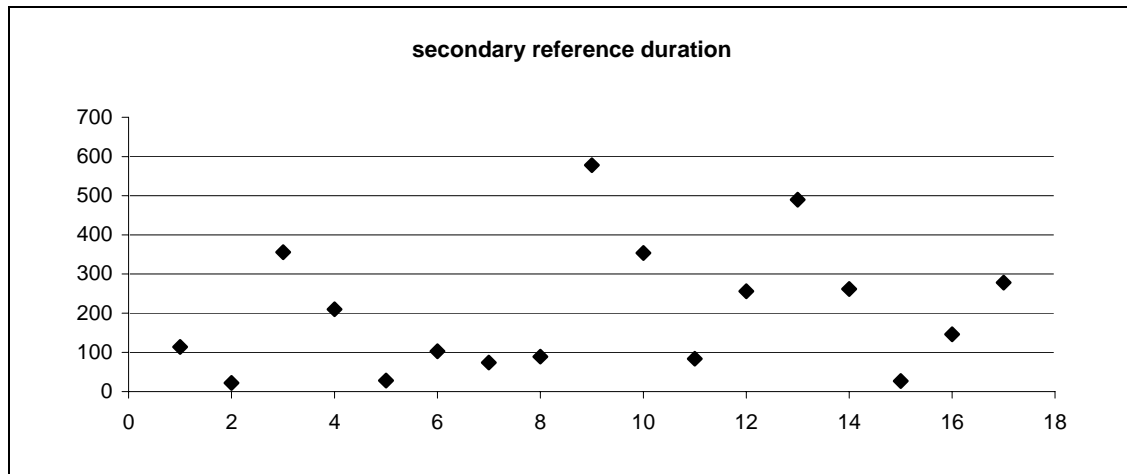


Fig. 104 Distribution of secondary reference responses according to their duration: 17 videos; mean 204”.

17 responses refer to CC or to CC’s video. However, rather than the topic of the video (as in the corresponding, commenting and spoofing ones), the reference is backgrounded, as in the case of an incidental mentioning. Here respondents use CC’s secondary reference in an interested way, as the chance to develop their own topic.

One of these responses features a pc screen with a *Word* document scrolling a text explaining ‘how to get to heaven’. The text, which ‘wants to spread the word of God’, introduces its author by referring to CC: ‘maybe you have seen some of my comments on ChrisCroker videos’ (Fig. 105). Here, the reference to the initiator is just used as an index of side-reference (a cohesive tie) so as to develop a completely different topic while preventing the response from being totally unrelated.

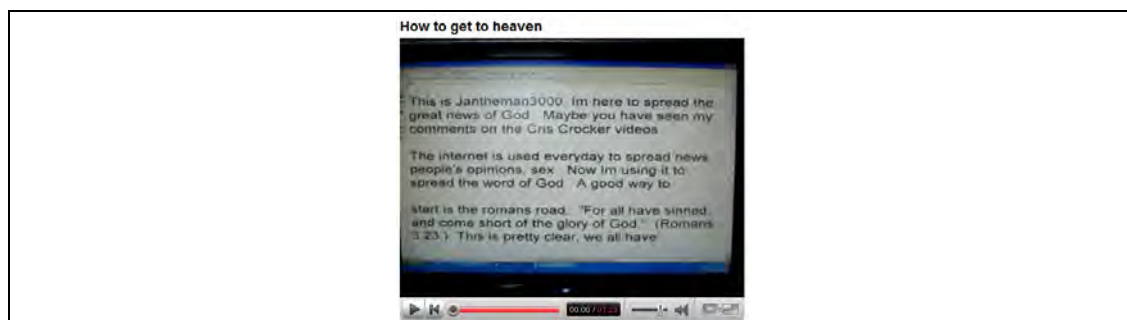


Fig. 105. The secondary reference to CC in a scrolling Word document response.

Some responses take the chance of referring to CC’s initial video or character to develop a discussion about videos on *YouTube* or to criticize the structure of the Website. For example, a response features a man asking the ‘YouTube community to stand up because now there’s no more funny videos, we need to wake up’. He talks about *YouTube* partnership programme (‘I’m not making a living out of it’); he



maintains to be gay and against gay-hate on *YouTube*, but all he

can't see is these assholes that can't see past their mirror. I don't have a problem with Chris Crocker being gay I have a problem with him being an asshole

After this unique reference to the thread initiator, he criticizes a series of genres of videos on *YouTube*, i.e. the ones he calls the 'movement videos' ('like get this person back off on youtube'), those staging puppets, the Christians vs. Atheists debate ('you are not going to convince each other, so why are you making videos?'). He criticizes further (You)Tube celebrities (e.g., *renetto*), the 'ring tones spam videos' and the 'puppy videos'. He concludes: 'stop uploading crap like this people... youtube community line up!'. Here again, CC's reference is used as the chance to develop a whole topic which concerns the (You)Tuber and deviates from the initial video's.

Another response, titled 'Youtube video's we can't stop making' (Fig. 106), starts with a typed 'Video's so good we can't stop making them', then it is structured through typed introductions to screenshots of videos. More precisely, a numbered genre label typed on a black screen introduces various shots of exemplary videos of that genre. Then the credits for each shot used in the video follow. Here, the side-reference to CC appears in the videos charted as no. 5: 'we can't stop making' is 'CHRIS CROCKER IMITATIONS'.

In sum, these respondents find a way to post a related response but to divert from the initial video topic. It is again a way of using to one's own interests the visibility of the initial video and the notoriousness of the character, as in the 'inferential' responses, but this time explicitly referring to it and thus establishing a cohesive tie.

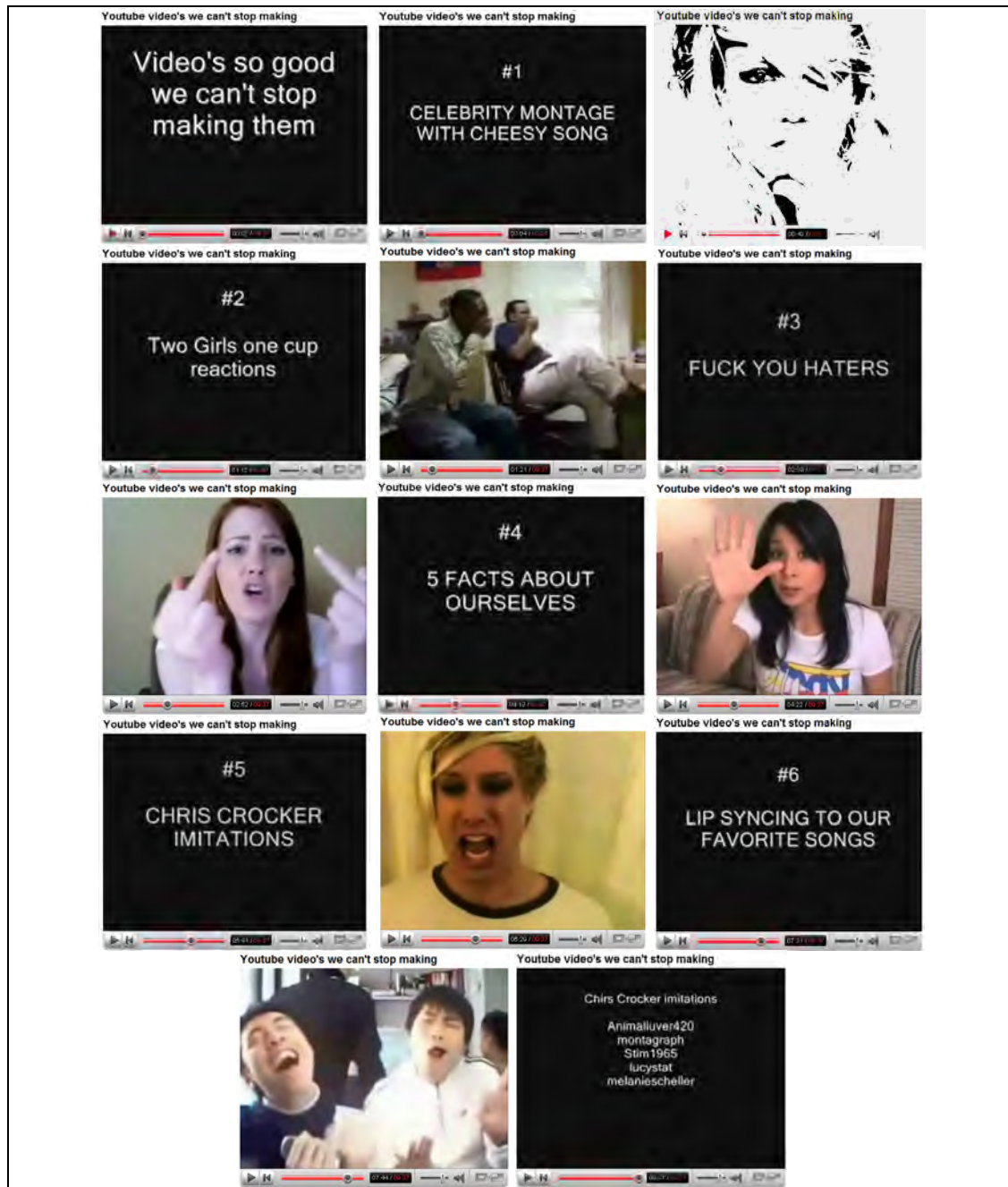


Fig. 106 CC side-reference included in a response which typifies popular *YouTube* video genres.

### 3.7 Random-related responses

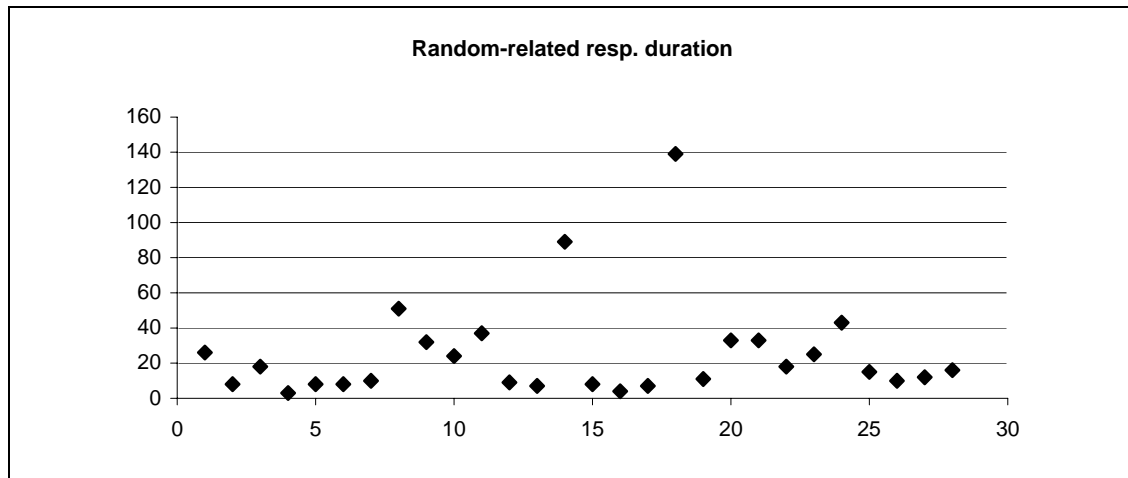


Fig. 107 Distribution of random-related responses according to their duration: 28 videos; mean 25”.

28 responses have nothing in common with the first video nor refer to elements of CC’s character or history. They are generally very short videos (cf. Fig. 107), which feature, e.g., a dog standing at a door shot from above, or a cat on a laptop. In a case the camera moves very fast from a boy’s naked chest up to the ceiling; another response films a white dog walking outdoor while a voiceover says ‘fluffy, come here fluffy’. A video features a long-shot of a woman standing at a seaside. One films a boy sitting on a sofa, while, behind him, two men play with a ball. Another video films a puppet representing the *Tweety*-bird cartoon character hanged on a stick outdoor, which suddenly explodes (Fig. 108); yet another one features a girl setting up her hair, wearing sunglasses and posturing while a voiceover behind the camera says ‘ready for the photo? One two three’, then the video ends.



Fig. 108. The exploding *Tweety*-bird response.

These are all videos whose ideational meaning seems not to communicate any ‘significant’ or ‘relevant’ message at all. Interpersonally, it is not clear why the uploader has made those videos and shared them online and what effect they are supposed to evoke in the viewer. On *YouTube*, these kind of videos correspond to a specific genre and are called ‘random’, i.e., videos which represent random things

without a precise purpose or ‘point’ (i.e., globally incoherent, in van Dijk’s terms; cf. Chapter 2, Section 2). When one considers that these videos were posted as responses to a video in which CC ‘does not do anything but blink’, an inference of relatedness can be made as to the ‘insignificance’, or ‘randomness’ of the content, which makes the video response another ‘Best video EVER!’. This relation is established by virtue of inferences, of a coherence which is sought for on the basis of a link of response between the two videos.

More specifically, since the ‘random’ video genre is well-established on *YouTube*, frequent viewers of *YouTube* videos are accustomed to encounter apparently meaningless videos<sup>124</sup>. Hence, rather than searching for a ‘significant’ meaning in the video (i.e., rather than assuming that the maxim of relevance is at work), they can take the randomness as a valid signified and thus consider it as the underlying topic of the video which relates it to the insignificance of the initial one. The ‘randomness’ thus becomes the actualized prompt-response relation of the exchange, the specific relatedness between the response and the initial video.

It must be noted that these ‘randomness’-related videos differ from the ‘insignificance’-corresponding responses discussed in 3.1, in that they do not present the same multimodal deployment of the initial video (i.e., a face staring the camera and enacting insignificant actions) and are therefore less attuned with (and thus less related to) the initial video.

### 3.8 Paratext-related responses

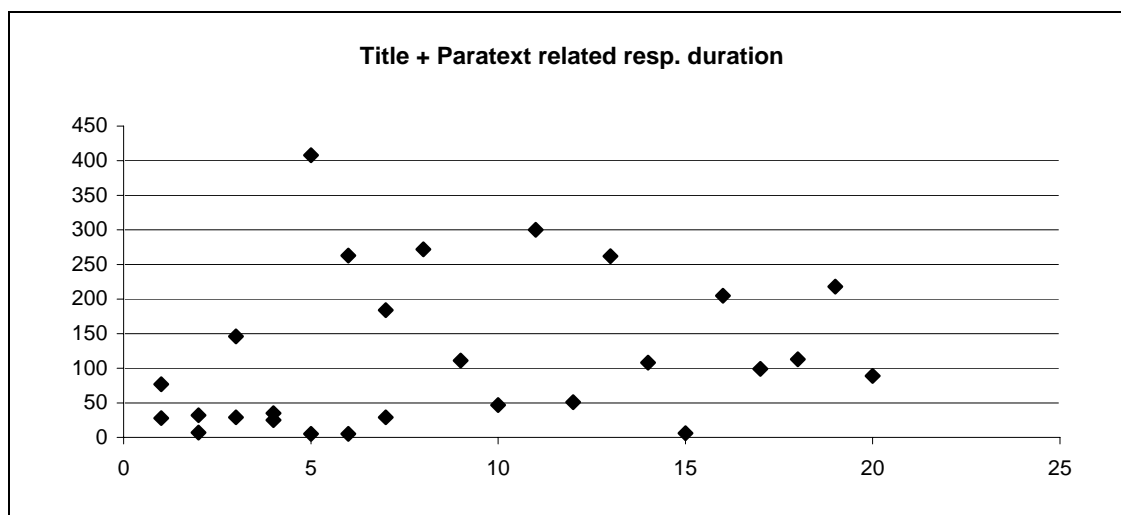


Fig. 109 Distribution of paratext-related responses according to their duration: 27 videos; mean 116”.

<sup>124</sup> The fact that *YouTube* hosts a huge number of home-made videos adds up to the viewers’ lack of expectations of ‘meaningful’ videos. Viewers know that (You)Tubers sometimes upload videos just for the sheer pleasure of doing it or just as a trial, i.e., to start practicing with the medium.

Another set of responses gives no clues to be related to any of the elements of the initial video, but is somewhat related to elements of its paratext.

In 20 responses, the initial video title, i.e., ‘Best Video EVER!’ is referred to, either in the response’s title (e.g., ‘the Best video ever 2’, ‘this is the best video of all time’) or somewhere else in the paratext (video description, e.g., ‘secound best video ever’, ‘THE GREATEST YOUTUBE VIDEO EVER’) or in the video itself (i.e., typed ‘this is the best video ever!’). These (You)Tubers interpret the initial video title according to their interest and take the ‘bestness’ value as a prompt to post what they consider their own – or somebody else’s – best video ever. A subset of these responses have a different, sometimes opposite, qualifying superlative in their title (i.e., ‘worst video ever’, or, ‘awsomest’, ‘THE WORST FILM EVER MADE!’, ‘THE ULTIMATE COMPILATION VIDEO’). These videos are totally unrelated to the initial video (so that they can be TV excerpts, music video clips, video game animations etc.), and the only clue of relatedness with the initial video is the cohesive tie (of either repetition or opposition) established by means of the qualifying superlative in the title, i.e., the (You)Tuber’s evaluation on the video itself.

Seven responses refer to the initial video description, in their video and/or in their description. They mainly refer to the views CC’s declared to achieve. So, for example, a response features a boy (concealed by b/w colour effects), wearing sunglasses, looking at the camera and saying:

hey what's going on *beantownbadboy* back again I just wanted to post a video response and get a point that you can get a ton of fucking views just by posting response to his video, you know what I'm saying I'm just like a complete fucking... but ehy, you got to watch this. peace

Moreover, the response description recalls the initial video’s one (i.e., ‘the point of this video is to prove...’): ‘Just to prove a point that if you post a response to Chris, you'll get a ton of hits too’. As in the ‘Where Do YouTube?’ thread (cf. Chapter 5, Section 3.3.4), these videos establish relatedness through cohesive ties (of repetition or opposition) in their (para)text.

### 3.9 Unrelated responses

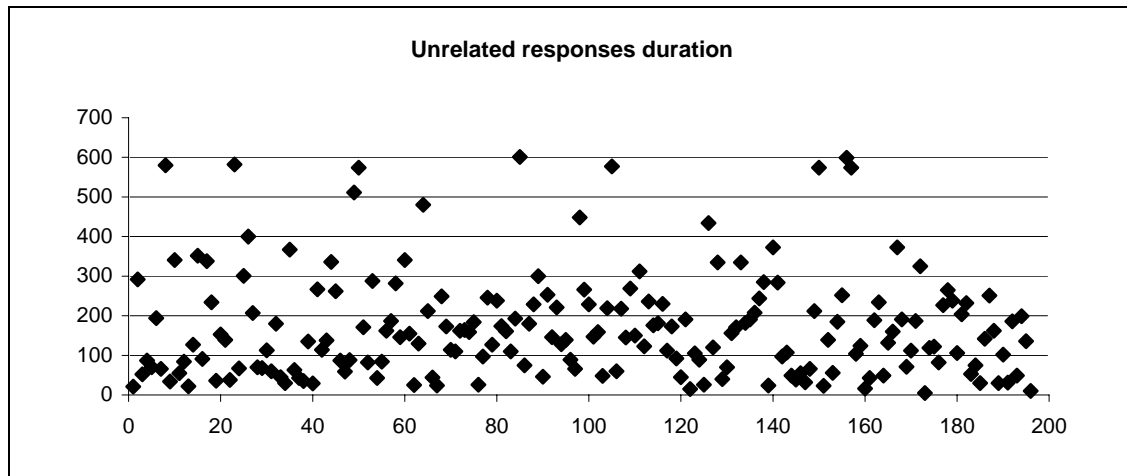


Fig. 110 Distribution of unrelated responses according to their duration: 196 videos; mean 171”.

Eventually, a considerable number of responses (196) give no clues to be related either to elements of the initial video (or to its paratext), or to elements of its uploader, i.e., CC’s persona and history. Arguably, by linking their videos as a response to a very popular one, these (You)Tubers aim at increasing the visibility of their videos, as also argued by a comment to one of these responses:

Stop posting your shit as video responces to popular videos to get yourself seen/ You suck!

One respondent declares explicitly this aim by typewriting on screen the following:

this is not a response to this featured video. I'm just abusing it to get views

A response receives an inquisitive comment concerning its relatedness to the initial video. It is replied by the response’s uploader, who, by answering with another – rhetorical – question, upholds the right of posting unrelated video responses:

Comment: Why is this a video response to Chris Crocker's video? This is totally a different topic...

Uploader’s reply: why did dinosaurs evolve from water?

While the comment expresses the expectation for a traditionally coherent exchange, the uploader’s reply supports the ongoing practice on *YouTube* of constructing interactional exchanges characterized by a breach of coherence and relevance. Significantly, the style of the reply is in itself a breach of these conventions, i.e., to the traditional rule of correctness according to which one cannot reply to a question with another question. Furthermore, the underlying discourse of the reply (i.e., the concepts of ‘evolutionary biology’ and ‘nature’ implied in dinosaurs evolving from water) re-connnotes the comment’s violation of what would be the communicative

‘norm’ – i.e., topic coherence – and assimilates it to a natural phenomenon, i.e., the taking up of an apt situation (i.e., a prompt) as the chance for developing (evolving) something completely different (unrelated), according to one’s interests.

By posting their unrelated responses to a popular video, these respondents contribute to enlarge the thread and make CC’s video charted among the most responded ones. Hence, even if they construct totally incoherent exchanges, these responses build successful interactions with the initial video, in that they concur to fulfil both the respondents’ and the initiator’s diversified interests. In other words, the (You)Tubers’ interests – i.e., linking one’s own videos to a popular one so as to enhance their hypertextuality – are compatible with CC’s (i.e., gaining visibility).

As in the analysis of the ‘Where Do YouTube?’ thread, also here, the unrelated category has been populated by means of a negative process, i.e., by considering the videos which cannot be placed in any of the other categories. Beside the fact that some (few) very short responses (cf. Fig. 110) could well be considered among the ‘insignificance’-related one, it must be noted also that another interpretation of relatedness with CC’s video is always possible for all these video responses, i.e., that each respondent has uploaded what she considers as the ‘Best video EVER!’. Indeed, some of these responses repost very famous *YouTube* video ‘memes’, like the ‘Hippo and Dog’ animation (Fig. 111.a.), the ‘Numa Numa’ dance, and the well-known ‘Bad Day At The Office’ video, shot from a CCTV camera, filming an office where a man, sat at his desk, grabs the pc monitor and destroys it in rage (Fig. 111.b.). These responses can be reasonably thought as having been re-posted by virtue of their popularity and thus because they are their uploaders’ judged ‘Best video EVER!’.

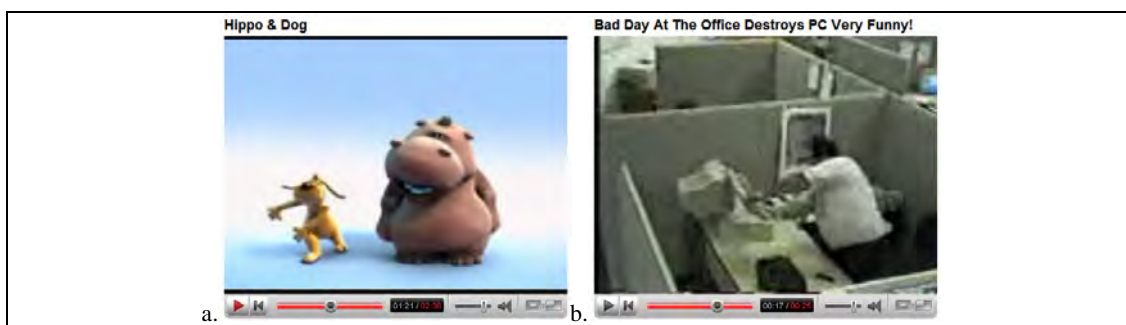


Fig. 111 Viral videos reposted as responses in the thread.



Fig. 112 An unrelated video judged as related by a comment.

At times, this interpretation is indeed made by the viewers, as exemplified by a comment to a totally unrelated response, titled ‘psychotic fear’, 29” long, featuring a man filmed outdoor sitting in his porch, shot from different angles, in black and white, with a suggestive soundtrack, then filmed in profile at a close shot, his eyes widened then closed as if in pain (Fig. 112). The comment reads as:

a lot better than CC

In other terms, these responses have been here considered as ‘unrelated’ only by virtue of the fact that no clues of relatedness are expressed either in the videos or in their paratexts. However, as evidenced by the aforesaid comment, independently from the respondent’s intentions (either to post a ‘bestness’ related response or to exploit an apt situation so as to get ‘dinosaurs evolve from water’), different interpretations of coherence and relatedness are possible. Different viewers, interested in different things and, possibly, coming across the video response from different paths (e.g., whether by following the ‘play all responses’ link from CC’s initial video or through keyword search) will perceive relatedness differently.

What is significant here is that both the uploader’s intentions and the interactants’ mutual understanding are irrelevant for interactional exchanges to be successful. Interactants exploit an affordance of the medium (i.e., the response option) to their purposes, thus constructing exchanges which can be (considered as) coherent to a greater or lesser extent. Totally incoherent exchanges can be fully acceptable if they fulfil the interactants’ interests, or rather, if the semiotic activity driven by the interactants’ interests – even if these are different – has compatible effects (e.g., making the initial video charted and increasing the response’s visibility on the Website). In their turn, viewers can interpret differently these exchanges, as more or less coherent exchanges or even as totally separate semiotic acts, in case they disregard the response link between them.

Video-interaction is a particularly loose form of communication, which does not bind tightly interactants to the meaning underlying their semiotic acts. It does not assign a great deal of responsibility to the authors for their (intended) meaning. As a consequence, viewers are given full freedom (or, in a different view, responsibility) of interpretation, both of the videos themselves (and the intentions with which they were posted) and of the interactional exchanges (i.e., of recovering or, in some cases, of building anew, the relatedness), stemming from the prompts given by the texts.

In sum, totally incoherent exchanges work successfully in video-interaction (unlike spam in email, which has given rise to the creation and the purchase of expensive anti-spam software tools)<sup>125</sup>, if they fulfil the interactants’ interests – which go

---

<sup>125</sup> This observation does not necessarily exclude that, in a near future, the interface be enhanced so as to filter unrelated responses. This again depends on the Website owners’ interested interpretation of the (You)Tubers’ practices. For a study which equals unrelated responses to spam, cf., Benevenuto



beyond the mutual understanding of their communicative intentions.

## 4 SUB-RESPONSES

44 responses to CC's 'Best Video EVER!' are responded in their turn by a total of 130 sub-responses (which collected further 11 sub-sub-responses). This considerable number of sub-responses, spread over several tens of responses in the thread, enables the analysis to verify the types of relatedness found in the first-level responses. Indeed, the 130 second-level responses posted to 44 different videos substantially confirm the typologies discussed in Section 3; this confutes any possible skewing of the data caused by idiosyncratic features (i.e., the initiator's lack of censorship over the thread, as the main cause of the wide variety of first-level responses).

Hence, sub-responses generally confirm the range of prompt-response relations found in the first-level responses. Some sub-responses correspond the corresponding first-level response (analogously to the 'Where Do YouTube' sub-responses which answer the topic question in their turn). Other responses develop the discussion topic (on CC or on gay-related issues) taken up by the first-level response; others parody or spoof either the initial video or their responded response (sometimes even corresponding it, if it is a spoof itself), while yet others take up an element of the response which does not relate to CC's initial video and make it salient by responding to it, so that they divert from the initial topic of the thread.

Six sub-responses posted to a first-level 'corresponding' response exemplify most cases. The first-level response is a video titled 'Even Bestest video EVER!!!', a drawn animation of a man raising his arms and blinking while a suspense soundtrack ends in screams of exultation. Its description imputes CC's celebrity (who 'is a GENIUS') to the viewers' lack of appreciation for quality.

Out of the six sub-responses, two correspond the video (and CC's initial video); indeed, in one the (You)Tuber looks at the camera with his fist up and his eyes wide open, then he blinks emphatically and puts his tongue out. In the other, titled 'Next to Last Best Video EVER!!!', the (You)Tuber blinks twice and widens his eyes.

Two sub-responses are commentaries. One threatens CC, by means of a very fast typing appearing and disappearing on a white screen: 'ChrisCrocker should die!' (its description reads as 'its tru I tells ya!'). Another sub-response, titled 'Black Bare', both criticizes CC and pleads the video it responds to, by featuring a black bear-masked (You)Tuber who affirms that his video is better than *Chris Crocker's*, insults CC and says 'I like your video... don't stop speaking out'. This latter utterance is clearly addressed to the responded (You)Tuber, rather than to CC.

---

et al. (2008b), which aims at detecting 'spammers', to overcome the admitted 'subjectivity' in defining 'spam' responses (i.e., it polices the uploaders' behaviour instead of their texts).

Finally, the two last sub-responses take up the first-level response's formal element (i.e., the semiotic mode) as a prompt to which to respond. Indeed, these responses tune in with the drawn animation of the first response, rather than with the ideational meaning (which corresponds CC's initial video). So, in one of these sub-responses, a drawn stick-figure raises and lowers the head in rhythm with 'What is love' soundtrack; in the other, titled 'Awesomest Bestest video EVER!!!' (which takes up also the title of the first-level response), a drawn animation of a man with a bloody knife cuts his head and blood bursts everywhere. These two responses take up an element of the video which is not related to the initial one and make it salient in their realization of the prompt-response relation, in a similar way as to when a topic is diverted in conversation (and in the 'inferential' first-level responses above). As mentioned when analysing the 'Where Do YouTube' thread (Chapter 5, Section 3.2), in exchanges strongly characterized by a phatic type of communication, form is content, so that the topic/theme (the '*fil rouge*' which links the two videos together) can be the form of the representation (i.e., the mode) rather than its content.

Other sub-responses develop the topic of the first-level response. The video dealing with CC's assessment on 9/11 (discussed in 3.2) is further responded by a video titled 'Chris Crocker 9/11' in which the videoblogger criticizes CC's for not caring about 9/11 and for caring only about Britney Spears. Thus, the sub-response endorses and reinforces the first-level response's argument against CC in relation to 9/11.

Parodies can be corresponded or commented. The 'Leave Britney Alone' original spoof titled 'Leave the Leave Britney Alone Guy Alone' (discussed in 3.4) has six sub-responses, two of which are original spoofs in their turn (and, in this sense, they correspond – in genre – their responded video). In turn, the remix featuring a deformed (i.e., pregnant) CC (discussed in 3.3) is commented by a video of a girl facing the camera and saying that CC is 'just arching his stomach out' (it is debatable whether the statement is ironic or is seriously defending CC and thus indicative of the respondent's *ad litteram* interpretation of the remix) and by another commentary which discusses (negatively) CC's persona.

Often the inferential forum-discussion responses (3.5), particularly those raising gay-related issues, get responses, either supporting the argument of the responded video or against it. So, the first level response titled 'GOD HATES FAGS', featuring an elderly man claiming that the 'God of the Bible' hates homosexuals (3.5), is responded by five respondents who argue against his argument; one of them also refers to CC (while the first-level response never does) and to the harm that hate language produces to the whole community:

you people are discriminating Chris Crocker who I know he's been taken for a chick but its clear that he's gay he is willing to admit this to the world. When you post a reply to something like this to him and you put this shit on the idea is that you are creating a more stressful environment for people on YouTube

Not all sub-responses argue against the video they respond to. Two sub-responses

support the respondent and criticize CC in their turn. They are posted to a first-level commentary response which insults CC, i.e., a masked (You)Tuber saying among other things ‘let me advise you take your video camera throw it in a fire because you got no talent Chris Crocker’. Interestingly, one of the sub-responses addresses CC, while referring in third-person to the responded (You)Tuber:

Mr Crocker I think that you see that I’m not the only one. There have been plenty of other people including this guy who some say some like Foamy [the first-level respondent] that you know obviously are a little disturbed by the way you are expressing your homosexuality and I must admit that I did laugh at this video because for the mere fact that I can understand I can understand that guy frustration with the way you are acting about Britney Spears with the way you’re acting about with the way you’re acting about yourself with the way you’re acting.

By addressing the initiator instead of the first-level (You)Tuber, the respondent shows his understanding of sub-level responses as part of the thread (although they are not displayed in this sense on the interface, cf. Chapter 4, Section 2.4.6).

Finally, as in the ‘Where Do YouTube’ thread (Chapter 5, Section 4.2), some sub-responses are self-responses, more or less related to their first-level response (i.e., topic/theme specification, development or diversion).

## 5 CONCLUSIONS

The analysis of the thread started by the ‘Best Video EVER!’ has firstly introduced and discussed the character and the semiotic activity of the initiator, *Chris Crocker*. Indeed, in reason of CC’s popularity achieved both inside and outside *YouTube*, the initiator’s character and (You)Tubing history are particularly significant to explain the initial video (and its main phatic function) and the type of thread built around it.

Secondly, the initial video has been described and analysed. Unlike the ‘Where Do YouTube?’ one, the ‘Best Video EVER!’ does not ask for responses but is rather a ‘prompting’ video, i.e., a video which ‘demands’ responses according to the playful provocative conventions embedded in the practice of video-interaction.

Thirdly, the analysis has focused on the responses, attempting a classification of the type of relatedness which they establish with the initial video. The following typology has been evidenced:

- a. corresponding responses;
- b. commentary responses;
- c. remakes;
- d. original spoofs;
- e. inferential responses;
- f. secondary-reference responses;

- g. random-related responses;
- h. paratext-related responses;
- i. unrelated responses.

As in the ‘Where Do YouTube’ thread, responses can variously (cor)respond to and attune with the initial video (a.) or can comment on it (b). They can further parody it by using its material (c.) or by enacting a spoof through the use of original material (d.), in both cases often framing the parody within other *YouTube* genres, so that part of the enjoyment resides in grasping the intertextual reference. Responses can also take up a background element of the initial video and make it salient, without referring explicitly to CC (e.), so that inferential processes are needed to retrieve the relatedness. They can further use an element of the initial video as a secondary reference (f.), thus taking the chance to deviate from it and develop their own topic. Responses can also take up a more abstract meaning of the video (its insignificance or ‘randomness’) and post other random material (g.). They can take up some elements of the initial video paratext (h.), i.e., the ‘bestness’ value of its title or the declared intention of the video description (to get views) and respond to it. Finally, a number of responses gives no clue to be related to the initial video (f.); these can be meant – without however expressing it explicitly – to be their uploaders’ ‘best video ever’ or, simply, can fulfil their uploaders’ interest in augmenting the hypertextuality of their videos, by linking them to a popular one.

All these differentiated prompt-response relations can be actualized either (or both) by taking up an element of the initial video or (and) an element of its uploader’s character and (You)Tubing history. In this latter case, the (You)Tuber’s sexual orientation, CC’s fanaticism for Britney Spears and CC’s famous past videos (particularly the ‘Leave Britney Alone’ one), together with CC’s achieved celebrity are the most frequently responded prompts. When they take up the initiator’s sexual orientation without referring to CC, responses can use the prompt ideationally, so that they start a hotly debated discussion forum on gay-related issues. In turn, others use it interpersonally and exploit the visibility of the thread as a meeting place for non-heterosexual participants.

Understandably, the less explicit the prompt-response relation, the more amount of shared knowledge and inferential work are required to perceive topic-relatedness. In fact, when a video is watched by virtue of its response link to another one, the viewer is prompted to retrieve any possible clues of relatedness. Furthermore, given that implicitness is widely practiced (implicit intertextual references are an obvious example of this), (You)Tubing conventions assume that a great deal of inferential work needs to be done, so that the Relevance Principle (i.e., achieving the maximum informational effect with the minimum inferential effort; cf. Chapter 2, Section 1.3) is not tenable in video-interaction.

Finally, the analysis of sub-responses in the thread has confirmed the range of prompt-response relations found in the first-level responses. This datum is significant

in confuting any possible skewing of the results on the first-level responses as a consequence of the initiator's celebrity status and lack of censorship on the thread.

In sum, the analysis o has highlighted three main functions of the video response option, which can be used to post:

- a reply related to the topic/form of the prompting text;
- a reply related to the (You)Tuber of the prompting text;
- a response prompted by a purpose embedded in the practice of (You)Tubing (e.g., to achieve visibility or to reach a certain target of viewers).

Video-interaction is used to reply on topic, to evaluate/judge/parody the initial (You)Tuber, or to reach a greater (or specific) audience, which is, understandably, a very frequent purpose of (You)Tubing.

Within these three general purposes, responses differentiate greatly themselves, both in their representations and in their interpersonal meanings (i.e., in what they express towards the initiator). Implicitness is greatly practiced, so that it is often unclear both whether the interpreted reference is the one intended by the respondent and whether the representation intends to praise or criticize the initiator.

In all cases, the interactants' mutual understanding of their intended meaning is not vital for successful exchanges to take place. The pleasure of transforming other people's texts in new ones, together with the challenge of exploiting the medium and its representational possibilities seem to be the main motives which drive video-interaction, so that any element, even (sometimes more crucially) the remotest one, can prompt a response. Ultimately, when the interactants' interests (rather than their communicative intentions) are compatible – even when different – successful communication does take place, no matter whether the participants understand each other, their semiotic acts or their underlying intentions.

What really seems to matter is the challenging pleasure of playing with the medium and representational resources. Hence, a great deal of the enjoyment in (both creating and interpreting) parodies, remakes and mashups, but also responses with implicit (intertextual) references, resides in the sign-maker's active process of meaning making, in the rewarding feeling of being an 'insider' by grasping the implicit meaning of the intertextuality, as a comment to a parody in the thread witnesses:

I saw the original!

Through the creative use of intertextual references, i.e., of selective salient elements from different sources transformed and recontextualized into new texts, *Tuberland* seems a light-hearted Eliot's *Waste Land* (1922), which enjoys the art of *pastiche* and plays with implicit references rooted in pop-culture (rather than, or more often

than, in high-culture and ancient myths).

In video-interaction, cooperation is newly shaped in the sense of an individual participation. Less importance is given to the mutual understanding of each other's intentions, in favour of an interested use of each other's acts as prompts for a creative transformation of texts which instantiate the participants' distinctiveness through the production of variations stemming from a given pattern.

Video-interaction epitomizes the contemporary values of agency and participation, conceived as individualized semiotic acts which produce selectively within chains of semiosis. Indeed, sign-making is produced through the interested selection, transformation, assemblage and recontextualization of signifiers prompted by other texts and acts. As in many forms of contemporary communication, sign-making produced through a copy-and-paste technique disregards traditional patterns of coherence and cohesion. This influences deeply 'acceptability' in text production and inserts within the profound changes in semiosis and communication of the present times.

This newly shaped form of participation lacks the traditionally conceived 'communalness', in favour of an enhanced individualization. So that communities are more of a discourse on rather than an apt metaphor of these agglomerated networks of sign-makers who link each other's semiotic acts according to their interests. In this sense, video-interactants can be rightly considered as emblems of the post-modern rhetor (cf. Chapter 2, Section 3.1).

## CHAPTER 7 CONCLUSIONS

*‘What does it matter who is speaking, someone said,  
what does it matter who is speaking’*

S. Beckett, *Texts for Nothing* (1954)

### 1 ACHIEVEMENTS OF THE RESEARCH

#### 1.1 Theoretical and empirical achievements

After introducing the object, aim and scope of the research in Chapter 1, I have reviewed in Chapter 2 the traditional coding-decoding and inferential models of communication (Grice, 1957, 1975; Shannon and Weaver, 1949; Sperber and Wilson, 1986), together with the notions of coherence and cohesion traditionally used in text analysis (Beaugrande and Dressler, 1981; Fairclough, 1992; Halliday and Hasan, 1976; van Dijk, 1985). I have confronted these theories and notions with the practices which take place in the interaction by means of videos on *YouTube*.

In spite of their differences, their common definition of successful communication on the basis of the interlocutors’ mutual understanding makes these theories and notions inapt for a thorough explanation and description of the sign-making practices of video-interaction. The analysis of both the process and the texts of video-interaction, in Chapters 4-6, have brought strong evidence to the criticism. This substantiated inadequacy of traditional models and notions to the description of video-interaction is, I believe, a theoretical achievement of the present research. Indeed, it has proved that contemporary forms of communication such as video-interaction function through selection, assemblage, transformation and recontextualization of previously existing texts (in a ‘copy-and-paste’ technique). As a consequence, they disregard coherence and cooperative (or relevance) principles, and rather work on the sign-maker’s interested response to the prompts offered by other texts, conceived as resources to produce new ones in chains of semiosis.

Stemming from the above, I have adapted the theoretical perspective of social semiotics (Hodge and Kress, 1988) and multimodal discourse analysis (Kress and van Leeuwen, 1996, 2006; Kress and van Leeuwen, 2001) to the analysis of the process and texts of video-interaction. The adaptation of this semiotic theory constitutes a further theoretical achievement of the research. Indeed, the study has developed the social semiotic notions of affordances and interest and has adapted them to interactional exchanges, by postulating a heuristic notion of interest-driven prompt-response relation, used here as the main tool of the analysis.

The analytical tool of interest-driven prompt-response relation has been applied to both the process and the texts of video-interaction. The analysis of the former in Chapter 4 has shown how the affordances of the medium – given by the distinctive features of the structure and by its representational possibilities and constraints – shape semiosis in a non-deterministic way. Indeed, affordances are exploited by interactants according to their interests, thus resulting in unexpected practices. In turn, these, when made socially dominant, can lead to changes in the structure itself. The analysis of the texts in Chapter 5 and 6 has shown that an individualized and interested participation shape a contemporary notion of cooperation, in which authorship is action in a semiotic space, rather than a certificate of ownership of the text or of its intended meaning in a community. This again constitutes, I believe, a theoretical achievement of the research, in developing an interpretative framework and a heuristic analytical tool for the description of ‘loose’ forms of contemporary communication such as video-interaction.

The main empirical achievement of the research lies in providing a description of a new form of communication, through the analysis carried out on both the process and the texts of video-interaction. Indeed, so far, no qualitative analysis has been carried out for its investigation, while the new and highly participated phenomenon of video-interaction has distinctive features which enable participants to communicate and interact through a unique combination of media and semiotic modes. Therefore, the research has started to fill a descriptive gap within our contemporary semiotic landscape.

The research has singled out and examined the distinctive features of the structural process of video-interaction, namely, (embodied and disembodied) multimodality, homogeneity and bidirectionality, publicity, asynchronicity, disembodiment, online communication, distance, multiple mediation and corporate interface distribution. The comparison of these distinctive features with their distribution in other forms of communication has enabled the mapping of video-interaction within our contemporary semiotic landscape. I have discussed the affordances derived from the distinctive features, as well as the technical and social ones of the video response option, and I have referred them to their uses in the interactants’ practices. Evidence has been brought to the fact that the interactants’ practices are greatly diversified, at times conflicting (following differentiated interests), and are developing incessantly, often leading to structural changes in the affordances; indeed, the video response option has been itself created on the interface as a result of the participants’ interested use of the already existing ‘related videos’ section. Over a 14-month period, I have observed the 40 most responded videos charted on the Website. The monitoring has led to a categorization of the types of videos which prompt the largest instances of video-interaction currently existing, identified in (1) video requests, (2) prompting videos, (3) anomalous instances, and (4) flooded-related responded videos. The 14-month monitoring period has also testified to ongoing changes in the observed practices, so that, as the top charted interactions have become larger and larger, the videos which start the largest exchanges have evolved from topic-



specificity towards genericity.

The so-obtained typology of most responded videos has provided the selective criteria for the video-threads included in the corpus. Carried out on almost two thousand videos, the analysis of the texts of video-interaction has evidenced a highly sophisticated complex of semiotic practices. Generally considered, representations strive to produce uniqueness by variously balancing differentiation and attuning with the other texts in the chain of semiosis. Patterns of variation-within-attuning are established through a differentiated use of a wide range of representational resources, through the exploitation of the semantic ambiguity/vagueness of the initial video and of the (remotest) formal elements prompted by it, as well as through the selection, assemblage and transformation of other texts, so that intertextual reference is highly practiced and implicitness characterizes frequently the referential system in the exchanges. Relatedness in the interaction can be established in various ways, from responding on, commenting, developing or deviating from the topic up to taking up a background element of the video and make it the salient responded prompt. The same can be done in relation to the initiator's character and (You)Tubing history, rather than on her video. Relatedness can be constructed also through formal attuning (rather than through semantic cohesion), i.e., by deploying the same salient mode of the initial video. Responses select any of the prompts offered by the initial video in an interested way, so that the actualized prompts are often taken as the chance to enact an attuned and unique performance. Mis-quotation and misinterpretation are generally acknowledged, and make manifest the fact that the interlocutors' intended meaning is disregarded for the establishment of successful communication in video-interaction. Incoherent exchanges (or traditionally considered 'marked' textual organizations) are frequently constructed and are totally accepted by the interactants. They are often the result of a recontextualization of previous texts in new exchanges (through 'copy and paste' techniques) according to the interactants' interests. Finally, also totally unrelated exchanges – or, rather, exchanges with no represented clues of relatedness – can be successful, when these comply with the interactants' diversified interests.

Rather than communicating something coherently and cooperatively, video-interaction works on a playful engagement with the medium and on an individualized and interested participation in the creative challenge of exploiting all semiotic resources and representational possibilities, prompted by other contributions in the exchange (i.e., stemming from a given kernel, either thematic or formal). In this sense video-interaction is similar to collective forms of artistic improvisation, which have illustrious antecedents in, e.g., the genre of variation in music.

Finally, the research has shown that the semiotic practices of video-interaction instantiate and epitomize the current changes in contemporary communication and semiosis (which, inevitably, reflect broader economic, political and social changes). For their thorough description, they need analogous changes in theories, models and analytical categories. We need an adequate perspective and suitable tools which can account for 'loose', individualized and participatory forms of communication based

on an interested selection, transformation, assemblage and recontextualization of resources. I believe that the attempt at doing it here constitutes the overall theoretical and empirical achievement of the research.

## 1.2 Methodological achievements

In Chapter 3 I have illustrated the research methodology. It has been devised so as to mitigate the shortcomings implied in the collection of data from *YouTube*, in terms of representativeness, significance, verifiability and reproducibility of the results. The popularity criterion adopted for the selection of the data has counterbalanced the aforesaid shortcomings; it has enabled the analysis to focus on the largest instances of video-interaction currently available and thus to observe the patterns of regularities and variations on very long chains of semiosis. This, I believe, is a methodological strength of the research. Having combined synchronous and short-term diachronical observation has also allowed the research both to grant the comparability of the texts and to detect the ongoing development of the practices.

Shaped by the research question and tailored by the specificity of the data, the adoption of an *ad hoc* transcription for the texts, driven by relevance, saliency and recurrence, has proved useful for handling a totally new type of materials in interaction. The so-devised transcriptive practice is another methodological achievement of the research, which could be used, developed and adapted for further studies on the subject.

In a cyclic process, the analytical methodology has involved all stages of the research, from data selection and transcription, to the analysis itself, up to the refinement of the theoretical framework and analytical tools. The adopted cyclic process has proved useful to investigate a new phenomenon. Indeed, given the unavailability of prior studies on the subject, the research design has required a series of progressive adjustments as the data were observed and new features emerged. The combination of quantitative and qualitative methods in the analysis focused on (selected) signifiers (i.e., on formal elements present in the texts) has corroborated with ‘facts’ the inevitably subjective interpretation of the data. I believe, it has proved useful to set the research free from the fear of ‘subjectivity’, which often refrains the analyst from drawing useful generalizations or courageous conclusions.

The covert observation conducted on the Website and the presentation of the data with no attempt at pseudonymizing them have granted the analysis the suitable environment, uninfluenced by the presence of the researcher. Ethically, these decisions have been motivated by the acknowledged publicity of the observed semiotic space and by my analytical focus on signifiers, on texts, rather than people. Furthermore, without denying the thorny ethical issue of disclosing sensitive information in research, in my view, video-interactants are to be regarded no less than film-makers and the ethical stance discussed in Chapter 3 should warn against

adopting any patronizing attitude towards them.

## 2 LIMITATIONS OF THE RESEARCH

I have focused the research on the processes, on the semiotic space and on the signifiers in video-interaction. On the one hand, it has enabled the description of a new form of communication, yet, on the other, it has inevitably constrained the scope of the study. Indeed, the major limitation of the research resides in not being able to say much of the intentions of the interactants, nor of their intended meanings, let alone their offline identities or practices, which would have required contacting and interviewing them. However, as discussed in Chapter 3, the recounts of their intentions and intended meaning would have constituted further representations, i.e., other texts to be interpreted and analysed as such. Even more, in the perspective of the present research, the elicited representations would have been the interested response to the prompts offered by the interviewer. Therefore, I have not tried to (pretend to) escape the circle of representation-interpretation by simply extending it to offline data or to research-driven ones. I have accepted this limitation of the study, and the inevitable possibility that the data may be interpreted differently by their producers or by other viewers, with the assumption that any interpretation always differs from another, even when done by the same person onto the same text at different times, in different contexts or to different interlocutors. Nevertheless, further ethnographic research on the subject could extend the scope of the analysis and compare the here-presented results with the representations given by the participants in interviews.

The focus of the research has necessarily constrained also the topics investigated. Indeed, even if the data provide useful material, the study has not dealt with issues such as identity construction, authenticity, discourses and authorship. Nevertheless, for the latter, my conception of author of a text as the one who uploads it (even when forwarding or reposting a pre-existing one) is a theoretical standpoint which brings Barthes' *Death of the Author* (1977a) a step further; and I believe that a conception of authorship as defined by action in a semiotic space, rather than by copyrights, is more attuned with contemporary communicative practices (cf., for example, wiki-based forms of text creation). I particularly regret that the research has only incidentally dealt with the production of humour in videos. Humour is maybe one of the most intriguing phenomena and plays a great role in making (often trivial) *YouTube* videos so enjoyable; yet, its thorough investigation would require a completely different research question.

The criterion of popularity adopted for the selection of the data has limited the observation to the largest instances of video-interaction, i.e., the threads started by the most responded videos on the Website. On the one hand, it has granted the corpus to be large enough to observe the patterns of regularity and variation of video-interaction. On the other, it has obviously disregarded the many small exchanges

occurring daily on *YouTube* by means of a few responses to videos posted among a close network of participants or, in all cases, by (You)Tubers who are not celebrities on the Website. These exchanges may not be that phatically marked and their contents may not be skewed by the ‘celebrity’ status of their initiators. The limitation has somehow been mitigated through the analysis of the sub-levels in the thread (i.e., of videos replying to respondents rather than to the popular video posted by the celebrity). The prompt-response relations actualized in the sub-responses have indeed confirmed the trends observed in the first-level responses. Furthermore, the focus on the patterns of relatedness among the videos, rather than on the contents of the videos themselves, has enabled the analysis to observe general trends which are rather independent of the specificity of the exchange. However, further investigation could extend the analysis to multiple small instances of video-interaction and compare the results with the present ones.

Finally, the ephemerality of video-interaction exchanges – and *YouTube*’s prohibition of storing online videos – has forcefully limited the analysis. Indeed, in its original design, I intended to analyse a further video-thread. Unfortunately it has disappeared from the Website during the monitoring period (its initiator’s account was hacked, cf. footnote 32); therefore its data could not be discussed in the research. The thread, started by a topic-generic video request, would have constituted the intermediate pole of a triad, from the topic-specific video request of the ‘Where Do YouTube?’ thread and the vague and phatic prompting video of the ‘Best Video EVER!’ thread. The digital incident has forced the thesis to limitate the discussion to the two extreme poles of ‘topic-specificity vs. genericity’. Nonetheless, they have given sufficient material for generalizations. Indeed, what observed as notable exceptions to coherence and cooperative principles in the topic-specific video-thread has proved to be the ‘rule’ in the phatically-characterized one, and can thus be assumed to be a distinctive feature of the patterns of relatedness in video-interaction.

### **3 FUTURE DEVELOPMENT**

Future research should broaden the corpus of texts and compare the results so far obtained. It could also diversify the criteria for the selection of data, so as to extend the analysis to a wider range of typologies of exchanges. Given the very recent birth of video-interaction and the continuous changes in the practices and affordances of the interface, a periodical monitoring of the Website could provide a useful follow-up to the developments of the semiotic practices of video-interaction as observed in the present research.

Furthermore, future research could verify the validity of the theoretical framework and analytical tools developed in the present research by applying them to other forms of communication, in particular to the ones employing digital technologies. I hope they will reveal useful to describe the changes in communication and representation of our times.

## REFERENCES

The following list includes only the works cited in the text.

- Adami, E. (2008a) 'Skills with the medium and available semiotic resources: the pattern of gaze in video-interaction'. Paper presented at the International Conference *Multimodality and Learning: New Perspectives on Knowledge, Representation and Communication*, Centre for Multimodal Research, London, UK, 19-20 June 2008.
- Adami, E. (2008b) 'Tubing the Web: a corpus-based study on video-communication'. Paper presented at *AACL 2008, American Association for Corpus Linguistics 2008*, Brigham Young University, Provo, Utah, USA, 13-15 March 2008.
- Adami, E. (2009a) "'Do YouTube?": When communication turns into video e-interaction'. In D. Torretta, M. Dossena and A. Sportelli (eds), *Forms of Migration – Migration of Forms: Atti del XXIII Convegno Nazionale AIA* Bari: Progedit.
- Adami, E. (2009b) "'To each reader his, their or her pronoun". Prescribed, proscribed and disregarded uses of generic pronouns in English', in A. Renouf and A. Kehoe (eds) *Corpus Linguistics: Refinements and Reassessments*.
- Adami, E. (2009c) "'We/YouTube": Exploring sign-making in video-interaction', *Visual Communication* 8 (4).
- Adami, E. (forth.) "'Where do YouTube?" Video-interaction: Differences communicated globally or global constraints on differences?' In R. Facchinetti, D. Crystal and B. Seidlhofer (eds.), *GlobEng, Global English. Cross-Cultural Perspectives* Bern: Peter Lang.
- Alibali, M. W., Flevares, L. and Goldin-Meadow, S. (1997). 'Assessing knowledge conveyed in gesture: Do teachers have the upper hand?' *Journal of Educational Psychology* 89: 183-193.
- Ariel, M. (2002) 'Privileged interactional interpretations', *Journal of Pragmatics* 34: 1003-1044.
- Atlas, J. (1989) *Philosophy without ambiguity*. Oxford: Clarendon Press.
- Atlas, J. (2005) *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and their Interface*. Oxford: Oxford University Press.
- Atlas, J. and Levinson, S. C. (1981) 'It-clefts, informativeness, and logical form: an introduction to radically radical pragmatics (revised standard version)', in P. Cole, *Radical*, 1-61. New York: Academic Press.
- Austin, J. L. (1962) *How to do things with words : the William James Lectures delivered at Harvard University in 1955*. Oxford: Clarendon Press.
- Bach, K. (1994a) 'Conversational implicature', *Mind and Language* 12: 124-162.
- Bach, K. (1994b) 'Semantic slack', in S. L. Tsohatzidis (ed) *Foundations of Speech Act Theory*, 268-291. London: Routledge.
- Bach, K. (1999a) 'The myth of conventional implicature', *Linguistics & Philosophy* 22 (327-366).

- Bach, K. (1999b) 'The semantics-pragmatics distinction: what it is and why it matters', in K. Turner (ed) *The Semantics/Pragmatics Interface from Different Points of View*, 65-84. Oxford: Elsevier.
- Bach, K. (2001a) 'Semantically speaking', in I. Kenesei and R. M. Harnish (eds) *Perspectives on Semantics, Pragmatics and Discourse*. A Festschrift for Ferenc Kiefer, 146-170. Amsterdam: John Benjamins.
- Bach, K. (2001b) 'You don't say?' *Synthese* 128: 15-44.
- Bach, K. (2002a) 'Seemingly semantic intuitions', in Campbell, J. K., O'Rourke, M., Shier, D. (eds) *Meaning and Truth*, 21-33. New York: Seven Bridges Press.
- Bach, K. (2002b) 'Semantic, pragmatic', in Campbell, J. K., O'Rourke, M., Shier, D. (eds) *Meaning and Truth*, 284-292. New York: Seven Bridges Press.
- Bakardjieva, M. and Feenberg, A. (2001) 'Involving the Virtual Subject: Conceptual, Methodological and Ethical Dimension', *Journal of Ethics and Information Technology* 2 (4): 233-248.
- Bakhtin, M. (1981) *The Dialogical Imagination*. Austin: University of Texas Press.
- Bakhtin, M. (1986) *Speech Genres and Other Late Essays*. Austin: University of Texas Press.
- Baldry, A. (2004) 'Phase and transition, type and instance: patterns in media texts as seen through a multimodal concordancer', in O'Halloran, K. L. (ed) *Multimodal Discourse Analysis. Systemic-Functional Perspectives*, 83-108. London/New York: Continuum.
- Baldry, A. and Thibault, P. J. (2006a) 'Multimodal cluster analysis approach', in Baldry, A. (ed) *Multimodality and Multimediality in the Distance Learning Age*, 24-31. Campobasso: Palladino.
- Baldry, A. and Thibault, P. J. (2006b) *Multimodal Transcription and Text Analysis. A Multimedia Toolkit and Coursebook*. London/Oakville: Equinox.
- Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., Ravichandran, D. and Aly, M. (2008) 'Video suggestion and discovery for youtube: taking random walks through the view graph', *Proceeding of the 17<sup>th</sup> international conference on World Wide Web* 895-904. Beijing, China: ACM New York, NY, USA.
- Bardzell, J. (2007) 'Creativity in amateur multimedia: popular culture, critical theory, and hci', *Human Technology, An Interdisciplinary Journal on Humans in ICT Environments* 3 (1): 12-33.
- Barnes, S. (2003) *Computer-Mediated Communication: Human to Human Communication across the Internet*. Boston: Allyn and Bacon.
- Barnes, S. and Hair, N. (2007) 'From banners to youtube: using the rear-view mirror to look at the future of internet advertising', *RIT Digital Media Library*: available at <http://hdl.handle.net/1850/7724>. Retrieved 10 February 2009.
- Barthes, R. (1977a) 'The Death of the Author', in Barthes, R. (trans. Heath, T. S.) *Image, Music, Text*, 142-148. New York: Hill and Wang.
- Barthes, R. (1977b) *Image-Music-Text*. London: Fontana.
- Barton, D. and Tusting, K. (2005) *Beyond Communities of Practice*. Cambridge: Cambridge University Press.

- Bateson, G. (1953) 'Why do Frenchmen...?' *ETC.: A Review of General Semantics* 10: 127-130.
- Bateson, G. (1972) *Steps to an Ecology of Mind*. Chicago: University of Chicago Press.
- Båve, A. (2008) 'A pragmatic defense of Millianism', *Philosophical Studies* 138 (2): 271-289.
- Baym, N. (2000) *Tune in Log On: Soaps, Fandom, and Online Community*. Thousand Oaks, CA and London: Sage.
- Beaugrande, R. and Dressler, W. (1981) *Introduction to Text Linguistics*. London: Longman.
- Beghtol, C. (2001) 'The Concept of Genre and Its Characteristics', *Bulletin of The American Society for Information Science and Technology* 27 (2): 17-19.
- Bell, D. (2001) *An Introduction to Cybercultures*. Oxon: Routledge.
- Benevenuto, F., Duarte, F., Rodrigues, T., Almeida, V., Almeida, J. and Ross, K. (2008a) 'Characterizing Video Responses in Social Networks', arXiv:0804.4865v1 [cs.MM] published online at: <http://aps.arxiv.org/abs/0804.4865>. Retrieved 11 May 2008.
- Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., Zhang, C. and Ross, K. (2008b) 'Identifying Video Spammers in Online Social Networks', in Castillo, C., Chellapilla, K. and Fetterly, D. (eds) *AIRWeb 2008, Fourth International Workshop on Adversarial Information Retrieval on the Web*, Beijing, China, April 22, 2008. *ACM International Conference Proceeding Series*, 45-52. New York: ACM.
- Benveniste, É. (1971) *Problems in General Linguistics*. Coral Gables: University of Miami Press.
- Benz, A. and van Rooij, R. (2007) 'Optimal assertions, and what they implicate. A uniform game theoretic approach', *Topoi* 26: 63-78.
- Berg, J. (1991) 'The relevant relevance', *Journal of Pragmatics* 16 (5): 411-425.
- Berlanga, A. J., Sloep, P. B., Brouns, F., Rosmalen, P. van, Bitter-Rijkema, M. E. and Koper, R. (2007) 'Functionality For Learning Networks: Lessons Learned From Social Web Applications'. *ePortfolio 2007 and HCSIT Proceedings (HR Technology, Digital Identity and Privacy conferences)*. Available at <http://hdl.handle.net/1820/1680> Retrieved 10 February 2009.
- Bezemer, J. and Kress, G. (2008) 'Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning', *Written Communication* 25: 166-195.
- Bezuidenhout, A. (2006) 'The coherence of contextualism', *Mind and Language* 21 (1): 1-10.
- Bezuidenhout, A. and Cooper Cutting, J. (2002) 'Literal meaning, minimal propositions, and pragmatic processing', *Journal of Pragmatics* 34: 433-456.
- Bhatia, V. (1993) *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Biber, D. (1989) 'A typology of English texts', *Linguistics* 27 (1): 3-34.
- Biber, D. and Finegan, E. (1986) 'An initial typology of text-types', in Aarts, J. and Meijs, W. (eds) *Corpus linguistics II*, 19-46. Amsterdam: Rodopi.

- Biber, D. and Finegan, E. (1994) *Sociolinguistic Perspectives on Register*. New York: Oxford University Press.
- Bird, G. (1979) 'Speech Acts and conversation—II', *The Philosophical Quarterly* 29 (115): 142-152.
- Blackmore, S. (1999) *The Meme Machine*. Oxford: Oxford University Press.
- Blakemore, D. (1987) *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (1995) 'Relevance Theory', in J. Verschueren, Ostman, J.-O. and Blommaert, J. (eds) *Handbook of Pragmatics: Manual*, 443-452. Amsterdam: John Benjamins.
- Blakemore, D. (2002) *Relevance and Linguistic Meaning*. Cambridge: Cambridge University Press.
- Blakemore, D. (2008) 'Apposition and affective communication', *Language and Literature* 17 (1): 37-57.
- Blass, R. (1990) *Relevance relations in discourse: A study with special reference to Sissala*. Cambridge: Cambridge University Press.
- Bourne, J. and Jewitt, C. (2003) 'Orchestrating debate: a multimodal approach to the study of the teaching of higher order literacy skills', *Reading: literacy and language, UKRA* July: 64-72.
- boyd, d. (2006) 'Friends, Friendsters, and Top 8: Writing community into being on social network sites', *First Monday* 11 (12).
- Breheny, R. (2006) 'Communication and folk psychology', *Mind & Language* 21 (1): 74-107.
- Brumark, Å. (2006) 'Non-observance of Gricean maxims in family dinner table conversation', *Journal of Pragmatics* 38: 1206-1238.
- Bruns, A., Wilson, J. A. and Saunders, B. (2007) 'Election Flops on YouTube. In ABC News Online: Club Bloggery, Australian Broadcasting Corporation', *Gatewatching*. Available at <http://eprints.qut.edu.au> Retrieved 10 February 2009.
- Burden, K. and Atkinson, S. (2007) 'Jumping on the YouTube bandwagon? Using digital video clips to develop personalised learning strategies', *ICT: Providing choices for learners and learning. Proceedings ascilite Singapore*. Available at <http://www.ascilite.org.au/conference/s/singapore07/procs/burden-poster.pdf> Retrieved 10 February 2009.
- Burgess, J. (forthcoming) "'All Your Chocolate Rain Are Belong to Us'? Viral Video, YouTube and the Dynamics of Participatory Culture', in Lovink, G. (ed) *VideoVortex collection*. Amsterdam: Institute of Network Cultures.
- Burgess, J. and Green, J. (2008) 'Agency and controversy in the YouTube community', Paper presented at *Internet Research 9.0: Rethinking Community, Rethinking Place* Copenhagen, 15-18 October, 2008. Available at <http://eprints.qut.edu.au/15383/> Retrieved 10 February 2009.
- Burgoon, J. K., Buller, D. B. , White, C. H. , Afifi, W. and Buslig, A. L. S. (1999) 'The Role of Conversational Involvement in Deceptive Interpersonal Interactions', *Personality and Social Psychology Bulletin* 25 (6): 669-686.
- Burt, S. M. (2002) 'Maxim confluence', *Journal of Pragmatics* 34: 993-1001.



- Cann, A. J. (2007) 'Podcasting is Dead. Long Live Video!' *Bioscience Education eJournal* 10. Available at: [www.bioscience.heacademy.ac.uk/journal/vol10/beej-10-C1.pdf](http://www.bioscience.heacademy.ac.uk/journal/vol10/beej-10-C1.pdf) Retrieved 10 February 2009.
- Capone, A. (2006) 'On Grice's circle (a theory-internal problem in linguistic theories of the Gricean type)', *Journal of Pragmatics* (38): 645-669.
- Capra, R. G., Lee, C. A., Marchionini, G., Russell, T., Shah, C. and Stutzman, F. (2008) 'Selection and context scoping for digital video collections: an investigation of youtube and blogs', *Proceedings of the 8<sup>th</sup> ACM/IEEE-CS joint conference on Digital libraries* 211-220. Pittsburgh PA, PA, USA: ACM, New York, NY, USA.
- Carlson, J., Heeschen, E. and Fatzinger-McShane, P. (2008) 'Communicating to Generation Y: Dietetic Interns Dissect You Tube Videos to Define What Is Necessary to Use It as a Communication Medium', *Journal of the American Dietetic Association*, 108 (9): 17-22.
- Carroll, S. (2008) 'The Practical Politics of Step-Stealing and Textual Poaching: YouTube, Audio-Visual Media and Contemporary Swing Dancers Online', *The International Journal of Research into New Media Technologies* 14 (2): 183-204.
- Carston, R. (1999a) 'Negation, 'presupposition' and metarepresentation: a response to Noel Burton-Roberts', *Journal of Linguistics* 35: 365-389.
- Carston, R. (1999b) 'The semantics/pragmatics distinction: a view from Relevance Theory', in Turner, K. (ed) *The Semantics/Pragmatics Interface from Different Points of View*, 85-125. Oxford: Elsevier.
- Carston, R. (2002a) 'Linguistic meaning, communicated meaning and cognitive pragmatics', *Mind & Language* 17: 127-148.
- Carston, R. (2002b) *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell Publishers.
- Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V. D.P., Loreto, V., Hotho, A., Grahl, M. and Stumme, G. (2007) 'Network Properties of Folksonomies', *AI Communications* 20 (4): 245-262.
- Cha, M., Kwaky, H., Rodriguez, P., Ahny, Y. and Moon, S. (2007) 'I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System', *IMC'07 Proceedings of the 7<sup>th</sup> ACM SIGCOMM conference on Internet measurement* San Diego, California, USA. 1-14. New York: ACM.
- Chandler, D. (1997) 'An Introduction to Genre Theory', Available at <http://www.aber.ac.uk/media/Documents/intgenre/intgenre.html> Retrieved 16 February 2009.
- Cheng, X., Dale, C. and Liu, J. (2007) 'Understanding the Characteristics of Internet Short Video Sharing: YouTube as a Case Study'. Published online at <http://arxiv.org/pdf/0707.3670>. Retrieved 10 February 2009.
- Chien, A. (2008) 'Scalar implicature and contrastive explanation', *Synthese* 161: 47-66.

- Chierchia, G. and McConnell-Ginet, S. (2000) *Meaning and Grammar*. Cambridge, MA: MIT Press.
- Claridge, C. (2007) 'Constructing a corpus from the web: message boards', in Hundt, M., Nesselhauf, N. and Biewer, C. (eds) *Corpus Linguistics and the Web*, 87-108. Amsterdam/New York: Rodopi.
- Clark, H. H. (1987) 'Relevance to what?' *Behavioral and Brain Sciences* 10: 714-715.
- Clemons, E. K., K., Barnett, S. and Appadurai, A. (2007) 'The Future of Advertising And the Value of Social Networks Websites: Some Preliminary Examinations', *ACM International Conference Proceeding Series; Vol. 258 Proceedings of the ninth international conference on Electronic commerce*. Minneapolis USA. 4 - 27 May 2007. 267-276. New York: ACM.
- Clift, R. (1999) 'Irony in conversation', *Language in Society* (28): 523-553.
- Cohen, M. A. and Küpçü, M. F. (2007) 'Congress and the "YouTube War"', *World Policy Journal* 23 (4): 49-54.
- Colston, H. L. (2000) "'Dewey defeats Truman" Interpreting ironic restatement', *Journal of Language and Social Psychology* 19 (1): 46-65.
- Cook, J., Pachler, N. and Bradley, C. (2008) 'Towards M-Maturity: The Nature and Role of Appropriation in Mobile Learning'. Paper presented at *mLearn 2008* Ironbridge Gorge World Heritage Site, Shropshire, UK, 8-10 October 2008.
- Cooren, F. and Sanders, R. E. (2002) 'Implicatures: a schematic approach', *Journal of Pragmatics* 34: 1045-1067.
- Costall, A. and Leudar, I. (2007) 'Getting over "the problem of other minds": Communication in context', *Infant Behavior & Development* 30: 289-295.
- Craig, R. and Tracy, K. (eds) (1983) *Conversational coherence*. Beverly Hills, CA: Sage.
- Crowston, K. and Williams, M. (2000) 'Reproduced and emergent genres of communication on the World-Wide Web', *The Information Society* 16: 201-216.
- Crystal, D. (1997) *English as a Global Language*. Cambridge: Cambridge University Press.
- Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.
- Crystal, D. and Davy, D. (1969) *Investigating English Style*. London: Longman.
- Davies, B. L. (2007) 'Grice's Cooperative Principle: Meaning and rationality', *Journal of Pragmatics* 39: 2308-2331.
- Davis, W. A. (1998) *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. Cambridge: Cambridge University Press.
- Davis, W. A. (2007) 'How normative is implicature', *Journal of Pragmatics* 39: 1655-1672.
- Dawkins, R. (1976) *The Selfish Gene*. Oxford: Oxford University Press.
- Debord, G. (1967) *La Société du spectacle*. Paris: Buchet-Chastel.
- Dedman, J. and Paul, J. (2006) *Videoblogging*. Indianapolis: Wiley Publishing.

- Delgrande, J. P., Nayak, A. C. and Pagnucco, M. (2005) 'Gricean Belief Change', *Studia Logica* 79: 97-113.
- Dicks, B., Soyianka, B. and Coffey, A. (2006) 'Multimodal ethnography', *Qualitative Research* 6 (1): 77-96.
- Duarte, F., Benevenuto, F., Almeida, V. and Almeida, J. (2007) 'Geographical Characterization of YouTube: a Latin American View'. *Proceedings of Web Conference, 2007. LA-WEB 2007. Latin American* 13-21. [ieeexplore.ieee.org](http://ieeexplore.ieee.org). Retrieved 13 March 2009.
- Duffy, P. (2007) 'Engaging the YouTube Google-Eyed Generation: Strategies for using Web 2.0 in Teaching and Learning', *The Electronic Journal of e-Learning* 6 (2): 119 - 130. Available at [www.ejel.org](http://www.ejel.org) Retrieved 10 February 2009.
- Eastment, D. (2007) 'Videos', *ELT Journal* 61 (1): 86-88.
- Eco, U. (1980) *Il nome della rosa*. Milano: Bompiani.
- Eichhorn, K. (2001) 'Sites unseen: Ethnographic research in a Textual Community', *International Journal of Qualitative Studies in Education* 14 (4): 565-578.
- Eisenstein, S. (1949) *Film Form: Essays in Film Theory*. New York: Hartcourt.
- Eisenstein, S. (1994) *Towards a Theory of Montage*. London: British Film Institute.
- Eliot, T. S. (1922) *The Waste Land*. New York: Boni and Liveright.
- Ess, C. (2001) 'AoIR Ethics Working Committee - a Preliminary Report'. Available online <http://aoir.org/reports/ethics.html>. Retrieved 14 January 2008.
- Eysenbach, G. and Till, J. E. (2001) 'Ethical Issues in Qualitative Research on Internet Communities', *BMJ* 323 (10 November 2001): 1103-1105.
- Facchinetti, R. (ed) (2007a) *Corpus Linguistics Twenty-five Years on*. Amsterdam: Rodopi.
- Facchinetti, R. (2007b) *Theoretical Description and Practical Applications of Linguistic Corpora*. Verona: QuiEdit.
- Fairclough, N. (1992) *Discourse and Social Change*. Cambridge: Polity Press.
- Fletcher, W. H. (2007) 'Concordancing the web: promise and problems, tolls and techniques', in Hundt, M., Nesselhauf, N. and Biewer, C. (eds) *Corpus Linguistics and the Web*, 25-45. Amsterdam/New York: Rodopi.
- Flewitt, R. (2005) 'Is every child's voice heard? Researching the different ways 3-year old children communicate and make meaning at home and in a pre-school playgroup', *Early Years* 25 (3): 207-222.
- Flewitt, R., Hample, R., Hauck, M. and Lancaster, L. (2009) 'What are multimodal data and transcription?' in Jewitt, C. (ed) *Handbook of Multimodal Analysis*. London: Routledge.
- Foucault, M. (1971) *L'ordre du Discours*. Paris: Gallimard.
- Fredsted, E. (1998) 'On semantic and pragmatic ambiguity', *Journal of Pragmatics* 30: 527-541.
- Freeman, L. (2006) *The Development of Social Network Analysis*. Vancouver: Empirical Press.
- Freitas, D., Buckenmeyer, J. and Hixon, E. (2008) 'YouTube.com for Teachers: A Useful Resource or Just More Hijinks?' in McFerrin, K. (ed) *Proceedings*

- of Society for Information Technology and Teacher Education International Conference 2008*, 4118-4119. Chesapeake, VA: AACE.
- Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.
- Gazdar, G. and Good, D. (1982) 'On a notion of relevance', in Smith, N. V. (ed) *Mutual knowledge*, 88-100. London: Academic Press.
- Gee, J. P. (2005) 'Semiotic social spaces and affinity spaces: from *The Age of Mythology* to today's schools', in Barton, D. and Tusting, K. (eds) *Beyond Communities of Practice: Language, Power and Social Context*, 214-232. Cambridge: Cambridge University Press.
- Gibbs, W. R. J. (1987) 'Mutual knowledge and the psychology of conversational inferences', *Journal of Pragmatics* 11 (5): 561-588.
- Gill, P., Arlitt, M., Li, Z. and Mahanti, A. (2007) 'Youtube traffic characterization: a view from the edge' *Proceedings of the 7<sup>th</sup> ACM SIGCOMM conference on Internet measurement* 15 - 28. San Diego, California, USA. New York: ACM.
- Giora, R. (1990) 'On the so-called evaluative material in informative texts', *Text* 10 (4): 299-320.
- Giora, R. (1993) 'On the function of analogies in informative texts', *Discourse Processes* 16: 591-611.
- Giora, R. (1995) 'On irony and negation', *Discourse Processes* 19: 239-264.
- Giora, R. (1997) 'Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory', *Journal of Pragmatics* 27: 17-34.
- Giora, R. (1999) 'On the priority of salient meanings: Studies of literal and figurative language', *Journal of Pragmatics* 31: 919-929.
- Giora, R., Meiran, N. and Oref, P. (1996) 'Identification of written discourse-topics by structure coherence and analogy strategies: General aspects and individual differences', *Journal of Pragmatics* 26: 455-474.
- Godwin-Jones, R. (2007) 'Emerging Technologies: Digital Video Update: YouTube, Flash, High-Definition', *Language Learning & Technology* 11 (1): 16-21.
- Goffman, E. (1963) *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press.
- Goodwin, C. (2001) 'Practices of seeing visual analysis: an ethnomethodological approach', in van Leeuwen, T. and Jewitt, C. (eds) *Handbook of Visual Analysis*, 157-182. London/Thousand Oaks: Sage.
- Goodwin, C. (2007) 'Participation, stance and affect in the organization of activities', *Discourse & Society* 18 (1): 53-73.
- Görlach, M. (2004) *Text Types and the History of English*. Berlin/New York: Mouton de Gruyter.
- Gray, J. (2004) *Consciousness: Creeping up on the Hard Problem*. Oxford: Oxford University Press.
- Greenbaum, S. (1991) 'ICE: the International Corpus of English', *English Today* 28: 3-7.
- Grice, H. P. (1957) 'Meaning', *Philosophical Review* 66: 377-388.
- Grice, H. P. (1967) 'Logic and conversation', *William James Lectures*, unpublished.

- Grice, H. P. (1975) 'Logic and conversation', in Cole, P. and Morgan, J. L. (eds) *Syntax and Semantics*, 41-58. New York: Academic Press.
- Gromik, N. (2007) 'From Film To Video Blogging: What Are The Steps?' in M. Thomas (ed) *e-Proceedings of "Wireless Ready Symposium: Podcasting Education and Mobile Assisted Language Learning"* 73-82. NUCB Graduate School Nagoya, Japan, 24<sup>th</sup> March 2007, available at <http://wirelessready.nucba.ac.jp/Gromik.pdf> Retrieved 10 February 2009.
- Gross, R. and Acquisti, A. (2005) 'Information revelation and privacy in online social networks', *Proceedings of WPES'05* Alexandria, VA, 71-80. New York: ACM.
- Gueorguieva, V. (2008) 'Voters, MySpace, and YouTube: The Impact of Alternative Communication Channels on the 2006 Election Cycle and Beyond', *Social Science Computer Review* 26 (3): 288-300.
- Haigh, C. and Jones, N. (2005) 'An Overview of the Ethics of Cyber-Space Research and the Implication for Nurse Educators', *Nurse Education Today* 25: 3-8.
- Halliday, M. A. K. (1978) *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.
- Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. Harlow: Longman.
- Halliday, M. A. K. and Hasan, R. (1985) *Language context and text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- Halvey, M. and Keane, M. T. (2007) 'Exploring Social Dynamics in Online Media Sharing, Poster Paper' *International World Wide Web Conference Proceedings of the 16<sup>th</sup> international conference on World Wide Web* Banff, Alberta, Canada, May 8-12, 2007, 1273-1274. New York: ACM.
- Hamman, R. B. (2001) 'Computer Networks Linking Network Communities', in Werry, C. and Mowbray, M. (eds) *Online Communities commerce, action, and the virtual university*, 71-95. New Jersey: Prentice-Hall.
- Hampel, R. and Hauck, M. (2006) 'computer-mediated language learning: Making meaning in multimodal virtual learning spaces', *The JALT CALL Journal* 2 (2): 3-18.
- Harnish, R. M. (1976) 'Logical form and implicature', in Bever, T. G., Katz, J.J. and Langedoen, T. (eds) *An Integrated Theory of Linguistic Ability*, 313-392. New York: Thomas Y. Crowell.
- Haugh, M. (2002) 'The intuitive basis of implicature: relevance theoretic implicitness versus Gricean implying', *Pragmatics* 12: 117-134.
- Heath, C. and Luff, P. (2007) 'Gesture and institutional interaction: figuring bids in auctions of fine art and antiques', *Gesture* 7 (2): 215-240.
- Heritage, J. (1984) *Garfinkel and ethnomethodology*. Cambridge: Polity Press.
- Herring, S. C. (ed) (1996) *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*. Amsterdam: Benjamins.
- Herring, S. C. (2001) 'Computer-Mediated Discourse', in Schiffrin, D., Tannen, D. and Hamilton, H. E. (eds) *The Handbook of Discourse Analysis*, 612-634. Malden/Oxford: Blackwell.

- Hindmarsh, J. and Pilnick, A. (2007) 'Knowing bodies at work: Embodiment and ephemeral teamwork in anaesthesia', *Organization Studies* 28 (9): 1395-1416.
- Hodge, R. and Kress, G. (1988) *Social Semiotics*. Cambridge: Polity.
- Hoey, M. (1983) *On the Surface of Discourse*. London: Allen and Unwin.
- Hoffmann, S. (2007) 'From web page to mega-corpus: the CNN transcripts', in Hundt, M., Nesselhauf, N. and Biewer, C. (eds) *Corpus Linguistics and the Web*, 69-85. Amsterdam/New York: Rodopi.
- Holt, R. (2004) *Dialogue on the Internet: Language, Civic Identity and Computer-Mediated Communication*. London: Praeger.
- Horn, L. (1985) 'Metalinguistic negation and pragmatic ambiguity', *Language* (61): 121-174.
- Horn, L. (2004) 'Implicature', in Horn, L. and G. Wards (eds) *The handbook of pragmatics*, 3-28. Oxford: Blackwell Publishers.
- Hundt, M., Nesselhauf, N. and Biewer, C. (eds) (2007) *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi.
- Jain, R. (2007) 'Photo Retrieval: Multimedia's Chance to Solve a Real Problem for Real People', *IEEE Published by the IEEE Computer Society* July-September 2007: 111-112.
- Jenkins, H. (2007) 'From YouTube to YouNiversity: Learning and Playing in an Age of Participatory Culture', *International Journal of Communication* 1: 145-146.
- Jewitt, C. and Kress, G. (2003) 'A multimodal approach to research in education', in Goodman, S., Lillis, T., Maybin, J. and Mercer, N. (eds) *Language, Literacy and Education: A Reader*, 277-292. Stoke on Trent: Trentham Books in association with the Open University.
- Johnson, M. (1987) *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Jones, L. K. (1977) *Theme in English expository discourse*. Lake Bluff, IL: Jupiter Press.
- Jones, S. G. (ed) (1998) *Cybersociety 2.0. Revisiting Computer-Mediated Communication and Community*. Thousand Oaks: Sage.
- Jones, S. G. (2004) 'Ethics and Internet Studies', in Johns, M. D., Chen, S. S. and Hall, G. J. (eds) *Online Social Research: Methods, Issues, & Ethics*, 179-186. New York: Peter Lang.
- Kasher, A. (1991) 'Pragmatics and the modularity of the mind', in Davis, S. (ed) *Pragmatics. A Reader*, 567-595. Oxford: Oxford University Press.
- Keen, A. (2008) *The Cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. New York: Doubleday.
- Kendall, L. (2002) *Hanging Out in the Virtual Pub: Identity, Masculinities, and Relationships Online*. Davis: University of California Press.
- Kendon, A. (1967) 'Some functions of Gaze-Direction in Social Interaction', *Acta Psychologica* 26: 22-63.

- Kessler, C. (2007) 'Where were you when YouTube was born?' *The Journal of Brand Management* 14 (3): 207-210.
- King, S. A. (1996) 'Researching Internet Communities: Proposed Ethical Guidelines for the Reporting of Results', *The Information Society* 12 (2, June 1): 119-128.
- Knobel, M. (2002) 'Rants, Ratings and Representation: Issues of ethics, validity and reliability in researching online social practices'. Paper presented at the annual meeting of *The American Educational Research Association* New Orleans, 3 April 2002, available online at: <http://www.geocities.com/c.lankshear/ethics.html>. Retrieved 2 September 2007.
- Kress, G. (1993) 'Against arbitrariness: the social production of the sign as a foundational issue in critical discourse analysis', *Discourse and Society* 4 (2): 169-193.
- Kress, G. (2003) *Literacy in the New Media Age*. New York: Routledge.
- Kress, G. (2008) 'New Literacies, New Democracies', *A challenge paper*. Available at: [http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/bch\\_challenge\\_paper\\_democracies\\_gunther\\_kress.pdf](http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/bch_challenge_paper_democracies_gunther_kress.pdf) Retrieved 10 February 2009.
- Kress, G. (2009) 'Assessment in the perspective of a social semiotic theory of multimodal teaching and learning', in Cummings, J. and Wyatt-Smith, C. (eds) *Educational Assessment in the 21<sup>st</sup> Century*, Chapter 2. New York: Springer.
- Kress, G. and Adami, E. (2009) 'A social semiotic analysis of mobile devices: Interrelations of technology and social habitus', in Pachler, N., Bachmair, B., Cook, J. and Kress, G. (eds) *Mobile learning outside and inside: structure - agency - practices?* New York: Springer.
- Kress, G., Jewitt, C., Ogborn, J. and Tsatsarelis, C. (2001) *Multimodal Teaching and Learning: The Rhetorics of the Science Classroom*. London: Continuum Books.
- Kress, G. and Pachler, N. (2007) 'Thinking about the 'm' in m-learning', in Pachler, N. (ed) *Mobile learning: towards a research agenda*, 7-32. London: WLE Centre, Institute of Education.
- Kress, G. and van Leeuwen, T. (1996, 2006) *Reading Images. The Grammar of Visual Design*. 2<sup>nd</sup> Edition. London: Routledge.
- Kress, G. and van Leeuwen, T. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Kress, G. and van Leeuwen, T. (2002) 'Colour as a semiotic mode: notes for a grammar of colour', *Visual Communication* (1): 343-369.
- Kristeva, J. (1969/1980) *Desire in Language: A Semiotic Approach to Literature and Art*. Oxford: Blackwell.
- Lancaster, L. (2001) 'Staring at the page: the function of gaze in a young child's interpretation of symbolic forms', *Journal of Early Childhood Literacy* 1 (2): 131-152.
- Lancaster, L. and Roberts, M. (2007) 'Marking on purpose', *Early Years Educator* 8 (11): 12-14.

- Lange, P. (2007a) 'Commenting on comments: Investigating responses to antagonism on YouTube'. Paper presented at *the Annual Conference of the Society for Applied Anthropology*, Tampa, Florida, March, 31 2007.
- Lange, P. (2007b) 'Searching for the "You" in "YouTube": An analysis of online response ability', *EPIC 2007, Ethnographic Praxis in Industry Conference*, 3-6 October, 2007 36-50. Keystone, CO USA: the American Anthropological Association.
- Lange, P. (2008) 'Publicly Private and Privately Public: Social Networking on YouTube', *Journal of Computer-Mediated Communication* 13 (1): 361–380.
- Lee, D. Y. W. (2001) 'Genres, Registers, Text Types, Domains, and Styles: clarifying the Concepts and navigating a Path through the BNC Jungle', *Language Learning & Technology* 3 (5): 37-72.
- Lee, E. (2008) 'Warming Up to User-Generated Content', *University of Illinois Law Review* 2008 (5). Available at SSRN: <http://ssrn.com/abstract=1116671> Retrieved 12 March 2009.
- Leech, G. (1983) *Principles of Pragmatics*. London/New York: Longman.
- Leech, G. (2007) 'New resources, or just better old ones? The Holy Grail of representativeness', in Hundt, M., Nesselhauf, N. and Biewer, C. (eds) *Corpus Linguistics and the Web*, 133-149. Amsterdam/New York: Rodopi.
- Lemke, J. (2006) 'Foreward', in Baldry, A. and Thibault, P. J. (eds) *Multimodal Transcription and Text Analysis. A Multimedia Toolkit and Coursebook*, xi. London/Oakville: Equinox.
- Levinson, S. C. (1983) *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. (1989) 'A review of Relevance', *Journal of Linguistics* 25: 455-472.
- Levinson, S. C. (2000) *Presumptive Meanings*. Cambridge, MA: MIT Press.
- Lewis, A. (2007) 'Online Social Networking: It's all Just Geek to Me', *Australian Counselling Association Journal* 7 (4). Available at [www.angelalewis.com.au/publ/Online\\_Social\\_Networking.pdf](http://www.angelalewis.com.au/publ/Online_Social_Networking.pdf) Retrieved 12 March 2009.
- Licklider, J. C. R. and Taylor, R. W. (1968) 'The computer as a communication device', *Science & Technology* 76: 21-31.
- Lindblom, K. (2001) 'Cooperating with Grice: a cross-disciplinary metaperspective on uses of Grice's cooperative principle', *Journal of Pragmatics* 33 (10): 1601-1623.
- Lüdeling, A., Evert, S. and Baroni, M. (2007) 'Using web data for linguistic purposes', in Hundt, M., Nesselhauf, N. and Biewer, C. (eds) *Corpus Linguistics and the Web*, 7-24. Amsterdam/New York: Rodopi.
- Machin, D. (2007) *Introduction to Multimodal Analysis*. London: Hodder Arnold.
- Machin, D. and Jaworski, A. (2006) 'Archive video footage in news: creating a likeness and index of the phenomenal world', *Visual Communication* 5: 345-366.
- Madden, M. (2007) *Online Video*. Washington, DC: Pew Internet & American Life Project.



- Malinowski, B. (1923) 'The Problem of Meaning in Primitive Languages', in Ogden, C. K. and Richards, I. A. (eds) *The Meaning of Meaning*, 146-152. London: Routledge.
- Mavers, D. (2007) 'Semiotic resourcefulness: A young child's email exchange as design', *Journal of Early Childhood Literacy* 7 (2): 153-174.
- McIntyre, G. (2003) 'A Sociological Examination of Positioning Strategies in an Online Postgraduate focus Group on Internet Research Ethics', MRes Dissertation, Institute of Education, University of London.
- McNeill, D. (1992) *Hand and Mind. What Gestures Reveal about Thought*. Chicago: Chicago University Press.
- Melican, J. and Faulkner, S. (2007) 'Getting Noticed, Showing-Off, Being Over-Heard: Amateurs, authors and artists inventing and reinventing themselves in online communities'. *EPIC 2007, Ethnographic Praxis in Industry Conference, 3-6 October, 2007* 51.65. Keystone, CO USA: the American Anthropological Association.
- Mey, J. L. and Talbot, M. (1988) 'Computation and the soul: A propos Dan Sperber and Deirdre Wilson's Relevance', *Journal of Pragmatics* 12 (5/6): 743-791.
- Miller, A. (1998) *Philosophy of Language*. London: UCL Press.
- Molyneaux, H., O'Donnell, S., Gibson, K. and Singer, J. (2008) 'Exploring the Gender Divide on YouTube: An Analysis of the Creation and Reception of Vlogs', *American Communication Journal* 10 (1). Available at [http://acjournal.org/holdings/vol10/01\\_Spring/articles/molyneaux\\_et.al.php](http://acjournal.org/holdings/vol10/01_Spring/articles/molyneaux_et.al.php) Retrieved 12 March 2009.
- Myers-Scotton, C. (1993) *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
- Neale, S. (1992) 'Paul Grice and the philosophy of language', *Linguistics and Philosophy* 15: 509-559.
- Norris, S. (2004) *Analyzing Multimodal Interaction: A methodological framework*. New York: Routledge.
- Norris, S. (2006) 'Multiparty interaction: a multimodal perspective on relevance', *Discourse Studies* 8 (3): 401-421.
- O'Brien, D. (2007) 'Viacom v YouTube and Google: copyright challenges for user generated intermediaries', draft paper prepared for the *ECUPL-QUT Legal and Policy Framework for the Digital Content Industry Conference, Shanghai 28-29 May 2007*.
- O'Donnell, S., Gibson, K., Milliken, M. and Singer, J. (2008) 'Reacting to YouTube Videos: Exploring Differences Among User Groups', *Proceedings of the International Communication Association Annual Conference (ICA 2008)* Montreal, Quebec, Canada. May 22-26, 2008. NRC 50361.
- Ochs, E. (1979) 'Transcription as Theory', in Ochs, E. and Schieffelin, B. (eds) *Developmental Pragmatics*, 43-72. New York: Academic Press.
- Pace, S. (2008) 'YouTube: an opportunity for consumer narrative analysis?' *Qualitative Market Research: An International Journal* 11 (2): 213 - 226.
- Paltridge, B. (1995) 'Working with genre: A pragmatic perspective', *Journal of Pragmatics* 23: 393-406.

- Paltridge, B. (1996) 'Genre, text type, and, and the language classroom', *ELT Journal* 50 (3): 237-243.
- Paltridge, B. (1997) *Genre, Frames and Writing in Research Settings*. Amsterdam: John Benjamins.
- Paolillo, J. C. (2008) 'Structure and Network in the YouTube Core', *Proceedings of the 41<sup>st</sup> Annual Hawaii International Conference on System Sciences* 156-164. IEEE Computer Society Washington, DC, USA.
- Récanati, F. (2002) 'Does linguistic communication rest on inference?' *Mind & Language* 17: 105-126.
- Récanati, F. (2004) *Literal Meaning*. Cambridge: Cambridge University Press.
- Regan, B. W. and Revels, A. (2007) 'Mapping the loss of reflexivity in the age of narcissism', *EPIC 2007, Ethnographic Praxis in Industry Conference, 3-6 October, 2007* 7-20. Keyston, CO USA: the American Anthropological Association.
- Reid, E. (1996) 'Informed Consent in the Study of Online Communities: A Reflection on the Effects of Computer-Mediated Social Research'. Available online <http://venus.soci.niu.edu/~jthomas/ethics/tis/go.libby>. Retrieved 15 January 2008.
- Roten, Y. d., Fivax-Depeursinge, E., Stern, D. J., Darwish, J. and Corboz-Warnery, A. (2000) 'Body and gaze formations and the communicational alliance in couple-therapist triads', *Psychotherapy Research* 10 (1): 30-46.
- Royce, A. (1984) *Movement and Meaning: Creativity and Interpretation in Ballet and Mime*. Bloomington: Indiana University Press.
- Sacks, H., Schlegoff, E. and Jefferson, G. (1974) 'A simplest systematics for the organization of turn-taking for conversation', *Language* 50: 696-735.
- Sadock, J. M. (1978) 'On testing for conversational implicature', in Cole, P. (ed) *Syntax and Semantics. 9. Pragmatics*, 281-297. New York: Academic Press.
- Sadock, J. M. (1986) 'Remarks on the paper by Deirdre Wilson and Dan Sperber', in Farley, A. M., Farley, P.T. and McCullough, K.-E. (eds) *Papers from the Parasession on Pragmatics and Grammatical Theory*. CLS 22, 85-90. Chicago, IL: Chicago Linguistic Society.
- Sampson, J. (1997) "'Genre," "style" and "register". Sources of confusion?' *Revue Belge de Philologie et d'Histoire* 75 (3): 699-708.
- Sandvig, C. (2006) 'The Internet at play: Child users of public Internet connections', *Journal of Computer-Mediated Communication* 11 (4): article 3.
- Saul, J. M. (2002) 'Speaker meaning, what is said, and what is implicated', *Noûs* 36: 228-248.
- Saussure, F. (1931) *Cours de Linguistique Générale* (ed. by Bally, C., Séchehaye, A. and Riedlinger, A.). Paris: Payot.
- Scott, J. (1991) *Social Network Analysis*. London: Sage.
- Searle, J. R. (1969) *Speech acts : an essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Sébastien, N., Ralambondrainy, T. and Rakotobe, M. (2007) 'A Simple Adaptive Learning Scenario For Web-Based Learning Environment: A Solution To

- The YouTube Issue?' *International Conference on Open and Online Learning, ICOOL'07* Penang, Malaysia, June 2007 Available at [www2.univ-reunion.fr/tralambo/files/sebastien2007learning.pdf](http://www2.univ-reunion.fr/tralambo/files/sebastien2007learning.pdf) Retrieved 10 February 2009.
- Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Sharf, B. F. (1999) 'Beyond Netiquette: The Ethics of Doing Naturalistic Discourse Research on the Internet', in Jones, S. (ed) *Doing Internet Research: Critical Issues and Methods for Examining the Net*, 243-246. Thousand Oaks: Sage.
- Sheehan, K. B. (2002) 'Toward a typology of Internet users and online privacy concerns', *The Information Society* 18 (1): 21-32.
- Shepard, M. and Watters, C. R. (1998) 'The evolution of cybergenres', *Proceedings of the 31<sup>st</sup> Annual Hawaii International Conference on System Sciences* 97-109. IEEE Computer Society Press.
- Shida, R. Y. and Gater, W. (2007) 'I Tune, You Tube, We Rule', *CAP* 1 (1): 30-31.
- Slevin, J. (2000) *The Internet and Society*. Cambridge: Polity.
- Smith, K. M. C. (2004) 'Electronic Eavesdropping: The Ethical Issues Involved in Conducting a Virtual Ethnography', in Johns, M. D., Chen, S. S. and Hall, G. J. (eds) *Online Social Research: Methods, Issues, & Ethics*, 223-238. New York: Peter Lang.
- Sperber, D. and Wilson, D. (1986) *Relevance. Communication and Cognition*. Oxford: Blackwell.
- Stainton, R. (1998) 'Quantifier phrases, meaningfulness "in isolation", and ellipsis', *Linguistics & Philosophy* 21: 311-340.
- Stalnaker, R. (1999) *Context and Content*. Oxford: Oxford University Press.
- Stam, R. (2000) *Film Theory: an introduction*. Oxford: Blackwell Publishers.
- Steen, G. (1999) 'Genres of discourse and the definition of literature', *Discourse Processes* 28: 109-120.
- Strawson, P. F. (1971) 'Meaning and truth', in Strawson, P. F. (ed) *Logico-linguistic Papers*, 13-45. London: Methuen.
- Streek, J. (1993) 'Gesture as communication I: Its coordination with gaze and speech', *Communication Monographs* 60 (4): 275-299.
- Sutton-Spence, R. and Woll, B. (1999) *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press.
- Swales, J. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Taylor, J. R. (1989) *Linguistic categorisation: Prototypes in linguistic theory*. Oxford: Clarendon.
- Thibault, P. J. (2000) 'The multimodal transcription of a television advertisement: Theory and practice', in Baldry, A. (ed) *Multimodality and Multimediality in the distance learning age*, 311-385. Campobasso: Palladino Editore.
- Thompson, G. and Hunston, S. (eds) (2005) *System and Corpus: Exploring connections*. London/New York: Equinox.

- Thurlow, C., Lengel, L. and Tomic, A. (2004) *Computer-Mediated Communication. Social Interaction and the Internet*. London: Sage.
- Tomasello, M. (2003) *Constructing a Language: a Usage-based Theory of Language Acquisition*. London: Harvard University Press.
- Tomlin, R. S. (ed) (1987) *Coherence and Grounding in Discourse*. Amsterdam: John Benjamins.
- Trier, J. (2007a) ““Cool” Engagements With YouTube: Part 2’, *Journal of Adolescent & Adult Literacy* 50 (7): 598-603.
- Trier, J. (2007b) ‘Media Literacy - “Cool” Engagements with YouTube: Part 1’, *International Reading Association* 50 (5): 408-412.
- Tseng, C. (2008) ‘Coherence and cohesive harmony in filmic text’, in Unsworth, L. (ed) *Multimodal Semiotics: Functional Analysis in Contexts of Education*, 87-104. London: Continuum.
- Turkheimer, M. (2007) ‘A YouTube Moment in Politics: An Analysis of the First Three Months of the 2008 Presidential Election’, *Report Student Work UEP Senior Comprehensive Projects, “Comps”*. Urban & Environmental Policy Institute Occidental College. Available at: [departments.oxy.edu/uepi/uep/studentwork/07comps/Turkheimer-Comps.pdf](http://departments.oxy.edu/uepi/uep/studentwork/07comps/Turkheimer-Comps.pdf) Retrieved 10 February 2009.
- Ulges, A., Schulze, C., Keysers, D. and Breuel, T. M. (2008) ‘A System That Learns to Tag Videos by Watching Youtube’, *Computer Vision Systems, Lecture Notes in Computer Science. 6<sup>th</sup> International Conference, ICVS 2008* Santorini, Greece, May 12-15, 2008, 415-424 Berlin/Heidelberg: Springer.
- van Dijk, T. A. (1972) *Some aspects of the grammars*. The Hague: Mouton.
- van Dijk, T. A. (1977) *Text and context*. London: Longman.
- van Dijk, T. A. (1979) ‘Relevance assignment in discourse comprehension’, *Discourse Processes* 2: 113-126.
- van Dijk, T. A. (1980a) *Macrostructures*. Hillsdale, NJ: Erlbaum.
- van Dijk, T. A. (1980b) ‘The semantics and pragmatics of functional coherence in discourse’, in Ferrara, A. (ed) *Speech act theory: Ten years later*, 49-65. Milano: Special issue of *Versus* 26/27.
- van Dijk, T. A. (1985) ‘Semantic discourse analysis’, in van Dijk, T. A. (ed) *Handbook of Discourse Analysis*, 103-137. London: Academic Press.
- van Leeuwen, T. (1999) *Speech, Music, Sound*. London: Macmillan.
- van Leeuwen, T. (2005) *Introducing Social Semiotics*. London: Routledge.
- van Leeuwen, T. and Jewitt, C. (eds) (2001) *Handbook of Visual Analysis*. London/Thousand Oaks: Sage.
- Verdi, M., Hodson, R., Weynand, D. and Craig, S. (2006) *Secrets of Videoblogging*. Berkeley: Peachpit Press.
- Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walther, J. B. (2002) ‘Research Ethics in Internet-Enabled Research: Human Subjects Issues and Methodological Myopia’, *Ethics and Information Technology* 4 (2): 205-216.

- Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Webb, M. (2007) 'Music analysis down the (You) tube? Exploring the potential of cross-media listening for the music classroom', *British Journal of Music Education* 24: 147-164.
- Weiss, S. (2007) *Online Social Networks and the Need for New Privacy Research in Information and Communication Technology*. Frankfurt am Main/Germany: Johann Wolfgang Goethe-Universität.
- Wellman, B. (1996) 'Are personal communities local? A Dumptarian reconsideration', *Social Networks* 18 (4): 347-354.
- Wellman, B. and Berkowitz, S. D. (eds) (1988) *Social Structures: A Network Approach*. Cambridge: Cambridge University Press.
- Wellman, B. and Gulia, M. (1999) 'Virtual communities as communities: Net surfers don't ride alone', in Smith, M. A. and Kollock, P. (eds) *Communities in Cyberspace*, 167-194. London: Routledge.
- Wenger, E. (1998) *Communities of Practice*. Cambridge: Cambridge University Press.
- Whiteman, N. (2007) 'The Establishment, Maintenance and Destabilisation of Fandom: A Study of two Online Communities and an Exploration of Issues Pertaining to Internet Research', PhD Thesis, Institute of Education, University of London.
- Wilde, O. (1891) *The Picture of Dorian Gray*. London/New York/Melbourne: Ward Lock & Co.
- Willett, R. (forthcoming) 'Parodic practices: Amateur spoofs on video sharing sites', in Buckingham, D. and Willett, R. (eds) *Camcorder Cultures: Media Technology and Everyday Creativity*. Basingstoke: Palgrave Macmillan.
- Wilson, D. (1999) 'Relevance and Relevance Theory', in Wilson, R. and Chierchia, G. (eds) *MITECS Encyclopedia of Cognitive Sciences*, 719-722. Cambridge, MA: MIT Press.
- Wilson, D. and Sperber, D. (2004) 'Relevance theory', in Horn, L. and G. Ward (eds) *Handbook of Pragmatics*, 607-632. Oxford: Blackwell.
- Xia, M., Huang, Y., Duan, W. and Whinston, A. B. (2007) 'Implicit Many-to-One Communication in Online Communities'. *Communities and Technologies 2007: Proceedings of the Third Communities and Technologies Conference, Michigan State University* 265-274. London: Springer.
- Yahia, S. A., Benedikt, M. and Bohannon, P. (2007) 'Challenges in Searching Online Communities', *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. 35-41.
- Zhang, Q. (1998) 'Fuzziness - vagueness - generality - ambiguity', *Journal of Pragmatics* 29: 13-31.
- Zink, M., Suh, K., Gu, Y. and Kurose, J. (2008) 'Watch Global, Cache Local: YouTube Network Traffic at a Campus Network - Measurements and Implications'. *Proceeding of the 15<sup>th</sup> SPIE/ACM Multimedia Computing and Networking (MMCN'08), 2008*. available at:

<http://gaia.cs.umass.edu/networks/papers/MMCN08-0.2.pdf>. Retrieved 10 February 2009.

Ziv, Y. (1988) 'On the rationality of 'relevance' and the relevance of "rationality", *Journal of Pragmatics* 12 (5/6): 535-545.